

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:27:40 ; Search time 17 Seconds

(without alignments)
1914.344 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 3313
Sequence: 1 MSSRAGPMSEKKNVGGYRP.....EFDEALEQLMKHLEHNEED 625

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	196	5.9	2130 1	BA2B_CHICK
2	190	5.7	1360 1	CING_XENIA
3	183	5.5	1191 1	CING_MOUSE
4	180	5.4	1972 1	BA2B_HUMAN
5	178	5.4	1197 1	CING_HUMAN
6	174	5.3	1937 1	MYH8_HUMAN
7	173.5	5.2	978 1	RA50_AQUAE
8	173	5.2	1084 1	MYSS_RABIT
9	172.5	5.2	593 1	TARA_HUMAN
10	172.5	5.2	1407 1	TRHY_RABIT
11	171.5	5.2	2116 1	MY52_DICDI
12	169	5.1	1938 1	MYH4_RABIT
13	167.5	5.1	1898 1	TRHY_HUMAN
14	166.5	5.0	1938 1	MYHD_HUMAN
15	165	5.0	724 1	HMFR_HUMAN
16	165	5.0	919 1	INCE_HUMAN
17	165	5.0	2663 1	CENE_HUMAN
18	163.5	4.9	1939 1	MYH1_HUMAN
19	163	4.9	1756 1	PEPL_HUMAN
20	163	4.9	1790 1	USO1_YEAST
21	162	4.9	966 1	STKA_MOUSE
22	162	4.9	1969 1	MYSA_CAEEL
23	161.5	4.9	1875 1	MLP1_YEAST
24	161.5	4.9	3911 1	AKA9_HUMAN
25	161	4.9	1102 1	MYSC_CHICK
26	160	4.8	1972 1	MYHB_HUMAN
27	159.5	4.8	1549 1	TRHY_SHEEP
28	159.5	4.8	1556 1	PROS_DROVI
29	159.5	4.8	1938 1	MYH6_RAT
30	159.5	4.8	2230 1	GOM4_HUMAN
31	158.5	4.8	1979 1	TRIA_HUMAN
32	158	4.8	1978 1	MYHB_CHICK
33	157.5	4.8	1941 1	MYH2_HUMAN

34	157.5	4.8	1972 1	MYHB_MOUSE	008638 mus musculus
35	156.5	4.7	976 1	SCPI_HUMAN	015431 homo sapien
36	156.5	4.7	1755 1	PEPL_MOUSE	092869 mus musculus
37	156	4.7	1169 1	SMC_MENTJA	059037 methanococc
38	155.5	4.7	1427 1	REST_HUMAN	030622 homo sapien
39	155	4.7	1020 1	CF60_HUMAN	08nd25 homo sapien
40	155	4.7	1085 1	YAPF_SCHPO	009863 schizosacch
41	154.5	4.7	1972 1	MYHB_RABIT	035748 oryctolagus
42	154.5	4.7	2363 1	SPCO_MOUSE	062261 mus musculus
43	154	4.6	852 1	RA50_THEMA	091x11 thermocoga
44	154	4.6	1939 1	MYH6_HUMAN	013533 homo sapien
45	154	4.6	1940 1	MYH3_CHICK	002565 gallus gall

ALIGNMENTS

```

RESULT 1
ID      BA2B_CHICK      STANDARD;      PRT; 2130 AA.
AC      Q9DEI3;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Bromodomain adjacent to zinc finger domain 2B (Extracellular matrix
        protein F22).
GN      BA2B.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxId=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Yoon H., Philip N.U.;
RT      "Cloning of a new extracellular matrix protein expressed in retina.";
RL      Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC      - FUNCTION: May play a role in transcriptional regulation.
CC      - SIMILARITY: Belongs to the WAF family.
CC      - SIMILARITY: Contains 1 bromodomain.
CC      - SIMILARITY: Contains 1 DDT domain.
CC      - SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC      - SIMILARITY: Contains 1 PHD-type zinc finger.
CC      - SIMILARITY: Contains 1 PHD-type zinc finger.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; AF224275; AAC36791.1; -.
DR      HSSP; Q92831; 1B91.
DR      InterPro; IPR001487; Bromodomain.
DR      InterPro; IPR004022; DDT dom.
DR      InterPro; IPR001739; Methyl-CpG_bind.
DR      InterPro; IPR001965; ZnF_PHD.
DR      Pfam; PF00439; bromodomain; 1.
DR      Pfam; PF02791; DDT; 1.
DR      Pfam; PF01429; MBD; 1.
DR      Pfam; PF00628; PHD; 1.
DR      PRINTS; PR00503; BROMODOMAIN.
DR      SMART; SM00297; BROMO; 1.
DR      SMART; SM00571; DDT; 1.
DR      SMART; SM00249; PHD; 1.
DR      PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR      PROSITE; PS50016; ZF_PHD_2; 1.
DR      PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR      PROSITE; PS50016; ZF_PHD_2; 1.
DR      Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW      Nuclear protein; DNA-binding.

```

```

FT DOMAIN 693 743 MBD.
FT DOMAIN 1010 1075 DDT.
FT ZN_PING 1895 1945 PHD-TYPE.
FT DOMAIN 2039 2109 BROMODOMAIN.
FT DOMAIN 9 46 SER-RICH.
FT DOMAIN 184 208 SER-RICH.
FT DOMAIN 212 219 POLY-GLU.
FT DOMAIN 544 610 ASP-RICH.
FT DOMAIN 797 984 COILED COIL (POTENTIAL).
FT DOMAIN 882 886 POLY-LYS.
FT DOMAIN 1219 1262 ASP-RICH.
FT DOMAIN 1909 1912 POLY-LEU.
SQ SEQUENCE 2130 AA; 236140 MW; 208C48FBDBA68F70 CRC64;

Query Match 5.9%; Score 196; DB 1; Length 2130;
Best Local Similarity 20.9%; Pred. No. 0.01;
Matches 144; Conservative 96; Mismatches 262; Indels 186; Gaps 28;

QY 19 RPEVEQLVGLAGTRLASQDDGEMVSKKKNKPKNTSG-----KTWVSONSNP--P 71
DB 374 RSKREQYQTPPAQLKKQESSKNLKVLSLSSKPTGSPAHOKLUTSLNNHSPFLT 433
QY 72 RAMGQOQGRG-----SNVSGRGNVSGNGNGKGIQANISGRRLS 115
DB 434 NALLGNHPNGVIGSVIOEVPALTTKQKQKINESVAIASSTPESLPVNLIS---ACG 489
QY 116 RKYDNNFVAPPVSRPPELEGW-----NMQARGSAQH-----TAVGEPPV----- 157
DB 490 KTTGNRTLVVPSYSPVLPFGSGKDKPVSNNAVNAVKTQHLCPAKLVVEGFGVDSAPFS 549
QY 158 -----EDVDNASEENDSDALDSDDDLA-----SDDYSDVS 191
DB 550 SKESDSDSDDDDDDEDEDDED-----DSDSDSSESSESSEDSDSDSDSDSDSDSDS 607
QY 192 QKSHGRKQKWKPFPGSLDSLSLQI--NEPQKQMTCPACQNGPAILWYNLHPL-- 247
DB 608 DDTDEGKPLKTKTGSSIKSSSGPAHSTPLNLQVAKTPSSAPSL--CEPTQPAVFL 666
QY 248 -----AHARTGARVRLHRLAEVLEKD-----LQMGASVTPCGE 284
DB 667 GTTPSTLTSSHGCSIKRRRVDERELRVLLEYQMRETRINFGRLQGEVAVAFAPCGK 726
QY 285 -----IYG-QWKGLEDKDYELVWPWYIIMTRLDKDDNKWJMGNGQELLEYFD 335
DB 727 KLRQYEVVGVQVCLL-----KEEEVY--FCIRAMEGRGRPPNP-----DQGH 769
QY 336 KYEALPARHSYQPGHGRGMSVLMFESSA--TGYLEARLHRLAEWGLDRIAMGQKSMF 393
DB 770 SREESRRRRKGRPPNVGSTETFDSDALRLQLQOIAFOAQIKLRLKLOKQOARA 829
QY 394 SGGVROLVGFPLATKODLIDFNHOSQKTRIKFELK--SYQEMVYKELROISEDN---OOL 448
DB 830 AKEKKQQAALMAAE-----KROKEQIKIMKQOEKIKRLOQIRMEKEILRAQOI 878
QY 449 NYFNKLISKQNAKAVLESELEIMSEKLRL-----TAEADRIYQRTQWQHEQNRREMDAH 504
DB 879 LEAKKKKKKEAANKLLAEKRIKEMKROQAVLKHQOLEHRRIDMERRRQOMML- 937
QY 505 DRFDMDIKQIHERDAKENFEMLOOQERAKVVGQOONINPSSNDCCRAAEVVSFTI 564
DB 938 -----MKAMEAKKAEK--EHLKQEKR-----DEKRLN----- 964
QY 565 EFQEKEMEEVEERMLIKQEKQEMDM 592
DB 965 --KERKLEORLIELEM-AKELKKQVEDM 989

```

```

DT 10-OCT-2003 (rel. 42, Last annotation update)
DB Cingulin.
OS Xenopus laevis (African clawed frog).
NC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=20082893; PubMed=10613913;
RA Cordenonci M., D'Attri F., Hammar E., Parry D.A.D., Kenrick-Jones J.,
RA Shore D., Citi S.,
RT "Cingulin contains globular and coiled-coil domains and interacts with
RT ZO-1, ZO-2, ZO-3 and myosin."
RL J. Cell Biol. 147:1569-1581(1999).
RN [2]
RP FUNCTION IN TIGHT JUNCTION BIOGENESIS.
RX MEDLINE=97029068; PubMed=8875080;
RA Cardellini P., Davanzo G., Citi S.;
RT "Tight junctions in early amphibian development: detection of
RT functional cingulin from the 2-cell stage and its localization at the
RT boundary of distinct membrane domains in dividing blastomeres in low
RT calcium."
RL Dev. Dyn. 207:104-113(1996).
RN [3]
RP FUNCTION IN TIGHT JUNCTION BIOGENESIS.
RX MEDLINE=20400099; PubMed=10940624;
RA Fesenko I., Kurth T., Sheth B., Fleming T.P., Citi S., Hausen P.;
RT "Tight junction biogenesis in the early Xenopus embryo."
RL Mech. Dev. 96:51-65(2000).
RN [4]
RP INTERACTION WITH OCCUDIN.
RX MEDLINE=99421641; PubMed=10491082;
RA Cordenonci M., Turco F., D'Attri F., Hammar E., Martinucci G.,
RA Megsfo F., Citi S.;
RT "Xenopus laevis occludin. Identification of in vitro phosphorylation
RT sites by protein kinase CK2 and association with cingulin."
RL Eur. J. Biochem. 264:374-384(1999).
RN [5]
RP INTERACTION WITH F-ACTIN.
RX MEDLINE=21538627; PubMed=11682052;
RA D'Attri F., Citi S.;
RT "Cingulin interacts with F-actin in vitro."
RL FEBS Lett. 507:21-24(2001).
RN [6]
RP INTERACTION WITH ZO-1.
RX MEDLINE=22140336; PubMed=12023291;
RA D'Attri F., Nadalutti F., Citi S.;
RT "Evidence for a functional interaction between cingulin and ZO-1 in
RT cultured cells."
RL J. Biol. Chem. 277:27757-27764(2002).
RN [7]
RP FUNCTION: Probably plays a role in the formation and regulation of
RN the tight junction (TJ) paracellular permeability barrier,
RN possibly by linking ZO proteins to the actomyosin cytoskeleton.
RN -1- FUNCTION: Parallel homodimer (By similarity). Binds ZO-1 and ZO-2
RN in vivo.
RN -1- SUBUNIT: Acts as an F-actin bundling protein in vitro.
RN directly.
RN -1- TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight
RN junctions of polarized epithelia and some endothelia.
RN -1- DEVELOPMENTAL STAGE: A maternally synthesized protein. Found in
RN the apical cortex in the fertilized egg, where it is associated
RN with cytoskeleton filaments, it is recruited to tight junctions
RN before ZO-1 and occludin. Nascent tight junctions are in place
RN by the two-cell stage.
RN -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but
RN does not abolish colocalization with ZO-1.
RN
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL Outstation
RN at the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way

```



```

QY 15 QCGYRPE-----VEQLVGLAGTR--LASQDDGGEWEVSKKONKPKNTSGKTW 63
DB 31 RSGRRPARDARASTYGVAVRVLGIGAGFPVVLNSENKGTDSFGVQIKGNRRGSPGA--- 87
QY 64 VSONSNPPRANGQOQOGRGNSVSGRGNVSGRNGNGI-----QANISG-----RG 111
DB 68 LSGDSLSPENPYSQVGRPATRSSTSDSEPKDHLNGLINSQSQASLTGLAFMSPNRS 147
QY 112 RALSKRYDNNFVAPPVVS-----RPPV-----EGGNWQARGSGAQTAVQ 152
DB 148 TSLLE-----LAPKTSINTIDTFLSSVSLINKPSQSG-----QVRGRTGRT--R 196
QY 153 EPPVEDVDVNAEENSDALDDSDDLAS-----DDYDSVQSQSGSRKONKWPFKF 207
DB 197 TLPHQCKRSQSLDRLPRTREERHQSANHTRTGTYDNHVSQSKNPSQOSP-FSGF 255
QY 208 FGSIDS-----LSLEQINER-----ORWHCPACONGPGLDWY----- 241
DB 256 SRSRQTDVWLQSFETRDPAWVQFKSTPPLADORE-----TAPPGADHVKATYGI 309
QY 242 -----NLHPL--LAHARTKG--ARRVKLHRLAY----- 267
DB 310 LREBSSEASVRRKVSIVLEQWPLGMVSPASTKALAGQELTRKMBELQKKLDEYVK 369
QY 268 -----LEKLOMRGASVIPCGRYIGOMKG-----LGEDE----- 296
DB 370 RQKLEPSRVGLEROLEEKAEECHRLQELERKKEVQOSSKELONMKLLGQEEGRHGL 429
QY 297 -----KDY-----EIVWPVWYIINTFLDDDDDKML 323
DB 430 ENQVKELOLKLKHSQPSGKESILLKDLDTRELLBELLEKORVEQLREHETALK 489
QY 324 GNGNOELLEFPKYKALAAHSYG--POGRHGSVLMFESSATGYLEAER-----LH 373
DB 490 GALKKEEVASHDEVEHVRLOVQRTQELRRSQODATQDHA--LELRQMSLVELQ 546
QY 374 RELAEWGLDRIAMGCRSMFSGVQVLYGFLATKQDLDIFMHSQCKTRLKFKLSYQEM 433
DB 547 RELLEETSEETGHW--QSWFQKKEEL--RATKQEL--LOLRMEKEEMEELGEEKNEV 597
QY 434 VKEELRQI--SEDOQOLNYPFKKL-----SKONKHA-----KYLESELEIM 472
DB 598 LQGLDLQARASTRTDQVELEKKEELRRTQELKELOAEQOQVEGRRHNVLEKQALAL 657
QY 473 SEKL--RRTAENRIRIORTKMOEQN-----REEM 501
DB 658 REERDRGRELEQONLOQTLQRLQDCEASAKAVASETEAMMLGORATVETTLRTQ 717
QY 502 DAHDF-----FMDSIKOIHBRDAKEENFEMLOOE- 533
DB 718 EENDEFRRILGLEQLKEARGLAEGEAVEARLRDKVHLEVEKQLEALNMAQEEBG 777
QY 534 -----RAKV-----GOQONINPSSNDCKRAEVSFTIEOE----- 568
DB 778 NLAALAKTALVRLDEAQRGLARLQGEQALNRALLEEGRQREALRSKLEBOKRLNLR 837
QY 569 -----KEMEERFEREMLIKQCKEKMDKKHHEITPLEKEF--DEALEQUMYKGLH 621
DB 838 TVDRINKLEIQIGDSSKLALQOLAQMEDYKARKEVADARQADMSAEKNSGGLS 897
QY 622 NEDD 625
DB 898 RLQD 901

```

RESULT 4
 BAZB HUMAN STANDARD; PRT; 1972 AA.
 ID BAZB HUMAN Q96EA1; Q96E08; Q9252; Q9Y4N8;
 AC Q9U1F8; Q96EA1; Q96E08; Q9252; Q9Y4N8;
 DT 28-FEB-2003 (rel. 41, Created)
 DT 28-FEB-2003 (rel. 41, Last sequence update)
 DT 15-MAR-2004 (rel. 43, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2B (hMALp4).

```

GN BAZB OR KIAA1476.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Testis;
RX MEDLINE=20130112; PubMed=1062543;
RT Jones M.H., Hamana N., Nezu J., Shitane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RA Ohara O., Nagase T., Kikuno R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 753-1972 FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [4]
RP SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).
RC TISSUE=Melanoma;
RA Anorgue W., Winkner U., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).
RC TISSUE=Skeletal muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
RA Altschuld S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Slatopolsky M., Soares M.B., Toshitsuki S., Adami R.D., Mullaly S.J.,
RA Brownstein M., Loggiano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Raha S.S., McMan P.J., McKernan K.J., Garcia A.M., Gay L.J., Hulik S.W.,
RA Bosak S.A., Morley K.C., Hale S., McKernan K.J., Garcia A.M., Gay L.J., Hulik S.W.,
RA Richards S., Morley K.C., Hale S., McKernan K.J., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,
RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.W.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
RA Butterfield J.S.N., Jones J.E., Jones S.J.M., Marra M.A.;
RA Schmechel A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).
RA Itagaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kikawa M., Watanabe M., Hiroaka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto K., Nakamatsu A., Nakamura Y., Nagahara K.,
RA Masuko Y., Kanehori K.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1524-1972 FROM N.A.
RC TISSUE=Testis;
RA Pousetka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISM1.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Comment-Experimental confirmation may be lacking for some
CC isoforms;

```


CC Name=1;
CC IsoId=Q9UIF8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9UIF8-2; Sequence=VSP_000553;
CC Name=Inferred from Ref.4;
CC Name=3;
CC IsoId=Q9UIF8-3; Sequence=VSP_000554;
CC Note=Inferred from Ref.5;
CC TISSUE SPECIFICITY: Expressed at varying levels in several
CC tissues, whereas a smaller transcript was expressed specifically
CC in testis.
CC -1- SIMILARITY: Belongs to the WAL family.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -1- SIMILARITY: Contains 1 DDT domain.
CC -1- SIMILARITY: Contains 1 methyl-binding (MBD) domain.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -1- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 731.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB032355; BAA89212.1; -
DR EMBL; AB040909; BAA96000.2; ALT_INIT.
DR EMBL; AL834381; CAD39044.1; ALT_INIT.
DR EMBL; BC012576; AAH12576.1; ALT_FRAME.
DR EMBL; AK027612; BAB5231.1; ALT_INIT.
DR EMBL; AL080173; CAB45759.1; -
DR PIR; T12495; T12495.
DR HSSP; Q92831; IB91.
DR GeneW; HGNC:963; BAZ2B.
DR MIM; 605683; -
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004023; DDT dom.
DR InterPro; IPR001739; Methyl_CpG_bind.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;
KW Nuclear protein; DNA-binding; Alternative splicing.
FT DOMAIN 549 601
FT DOMAIN 891 956
FT ZN_FING 1735 1785
FT DOMAIN 1881 1951
FT DOMAIN 45 69
FT DOMAIN 73 79
FT DOMAIN 399 470
FT DOMAIN 641 672
FT DOMAIN 687 706
FT DOMAIN 706 845
FT DOMAIN 1100 1143
FT DOMAIN 1138 1179
FT VARSPPLIC 437 534
FT VARSPPLIC 593 626

FT CONFLICT 137 137 S -> F (IN REF. 1).
FT CONFLICT 226 226 S -> L (IN REF. 1).
FT CONFLICT 627 627 G -> E (IN REF. 1).
FT CONFLICT 722 722 E -> K (IN REF. 5).
FT CONFLICT 790 799 RIKEKEMRO -> QKKKKKKKK (IN REF. 5).
FT CONFLICT 1128 1128 K -> Q (IN REF. 1).
FT CONFLICT 1183 1183 Q -> P (IN REF. 1).
FT CONFLICT 1195 1195 Q -> R (IN REF. 1).
FT CONFLICT 1453 1453 L -> S (IN REF. 3).
FT CONFLICT 1838 1838 K -> Q (IN REF. 1).
SQ SEQUENCE 1972 AA; 220709 MW; DAF105C927B91569 CRC64;

Query Match 5.4%; Score 180; DB 1; Length 1972;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 114; Conservative 83; Mismatches 189; Indels 158; Gaps 21;

QY 140 QARGSAQHTAVGPEPVEDVDVNASEENDSDALDDSDDLAS-----183
DB 394 QFRGTDSDIPSSKDSNDEDEDEDEDEDEDESDSQSESDNSSESDTEGSEER 453
QY 184 --DDYSDVQSKSHSRKQKMKFFGSL--DSLIEQINERQRMHCPCQNGPGATD 239
DB 454 DDDDKQDDSDSDTEGKTSKMLKNTTSVSPSMSTGHSPTRNLIAXAPGAPALC 513
QY 240 WYMLHPL-----AAATKGAARVKLHRELAEVLEKD-----LQMR 275
DB 514 SESQSPAFIQTSSSTLTSSPHSCTSRRRYVTDERELRIPIEYQMGRETIRNFCGRLOGE 573
QY 276 GASVIFCGEITYGQ-----KGLGEDEKD-----IYW---303
DB 574 VAYYAPCGKRLQYPPVVIKYLSHNGIMDISRDNFSAKIRVGDIFYEARDGPGMGWCLL 633
QY 304 -----PPWVIMNTRLDKQNDNDLGMQNGEELLEFDKY---BALAPHSYSGQHRGMS 355
DB 634 KEEDVIFRIAMEGRGRPPNP-----DRQAPREESPMRRKKGPPVNGNA 679
QY 356 VLMFESSA--TGYLEARLHRELAEMGLDRIAMGQKSMFSGGVRLYGLATKODLDF 413
DB 680 EPLDNDADKLRLQAOELARQAQIKLARKQKQKQAVAKAKKQQAIMAAE-----734
QY 414 NQHSQKTRIKPELK-SYQEMVVKELQISEDN---QQLNYPKNLSKQNKIAXVLEES 468
DB 735 -----KRKQEQIKIMKQOEKIRIQIIRMEKELRAQQLLEAKKKKEKAANAKLLEAE 788
QY 469 LEIMSEKLRRTAEDNRIYQRTQMGHEQREENDADRFPMDSIKQIHERDAKEENFEM 528
DB 789 KRKEKEMR-----QOAVLKH-QERERRRH---MMLMKAMEARKKAEEK--ER 833
QY 529 LQQQERAKVVGQOQONINPSSNDDCKRAEVSFFIEPOKEKEEVEEREMLIKQOEKK 588
DB 834 LKQEK-----DEKLN-----KERKLEQRLELEM-AKELKKEP 866
QY 589 MEDM 592
DB 867 NEDM 870

RESULT 5
CING_HUMAN STANDARD; PRT; 1197 AA.
ID CING_HUMAN
AC Q9P2M7; Q9NR25;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cingulin.
GN CGN OR KIAA1319.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroepithelium;

RX MEDLINE=20499514; PubMed=11042084;
 RA Citi S., D'Atti F., Parry D.A.D.;
 RT "Human and Xenopus cingulin share a modular organization of the
 RT coiled-coil rod domain: predictions for intra- and intermolecular
 RT assembly.";
 RL J. Struct. Biol. 131:135-145(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [3]
 RP INTERACTION WITH TJPI.
 RX MEDLINE=22140336; PubMed=12023291;
 RA D'Atti F., Nadalucci F., Citi S.;
 RT "Evidence for a functional interaction between cingulin and ZO-1 in
 RT cultured cells.";
 RL J. Biol. Chem. 277:27757-27764(2002).
 CC - FUNCTION: Probably plays a role in the formation and regulation of
 CC the tight junction (TJ) paracellular permeability barrier.
 CC - SUBUNIT: Homodimer (By similarity). Interacts with TJPI/ZO-1.
 CC - TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight
 CC junctions of polarized epithelia and some endothelia. Expressed in
 CC pancreas, kidney, liver and lung, but not in skeletal muscle,
 CC placenta, brain or heart.
 CC - DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but
 CC does not abolish colocalization with ZO-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF263462; AAF74498.1; ALT_INT.
 CC EMBL; AB037740; BAA92557.1; ALT_INT.
 CC Genew; HGNC:17429; CGN.
 DR GO: GO:0005923; C: tight junction; NAS.
 DR GO: GO:0003779; F: actin binding; ISS.
 DR GO: GO:0005515; F: protein binding; IPI.
 DR InterPro: IPR002928; F: Myosin tail.
 DR Pfam; PF01576; Myosin_tail_1.
 DR Tight junction; Coiled coil.
 FT DOMAIN 1 351 HEAD.
 FT DOMAIN 352 1154 COILED COIL (POTENTIAL).
 FT DOMAIN 1155 1197 TAIL.
 FT DOMAIN 363 836 GLU-RICH.
 FT SITE 42 56 ZIM.
 FT SITE 106 400 INTERACTS WITH ZO-2.
 SQ SEQUENCE 1197 AA; 136385 MW; 0C9375283AABAF3D CRC64;
 Query Match 5.4%; Score 178; DB 1; Length 1197;
 Best local similarity 19.0%; Pred. No. 0.046;
 Matches 173; Conservative 121; Mismatches 284; Indels 332; Gaps 40;
 QY 15 QGGRPE-----VQLVQGLAGTR--LASQDQGEWEVISKKNKPKNTSGKTW 63
 DB 31 RGRPRPADAPASTGYAVVVGILAGPFIYVINGEGKGSFGVQIGANDQG--ASGA-- 87
 QY 64 VSQNSNPRAWGGQOQGRGNSVGRGNVSGRGNGRGRT---QANISRGALSRKYD 119
 DB 88 LSSDLELPENPYQYKFPAPQSGSTSDBEQAYWNGTLRSHQASLHAGGVPDPSNRS 147
 QY 120 NNFV--APPVSRP-----PLE-----GGMNQARGSGNQHTAVQEPDVE 158
 DB 148 NSMLBLAPVAVSPGSTIDTAPLVSVDLINFKDSQLGQ---QANGRTGRTRM--LPPEQ 202

QY 159 DDVNASEENDSDALDSDDD-----LASDDYDSDVSQKSHGRKQNKWFKKFPGLSDS 213
 DB 203 RRRSKSLDSRLRPFDFEERERQSTNHTSTSTKYNHNGTSGKQPAQSQN-----LSP 253
 QY 214 LS-----TEQINPQRMWCPACQN-----GPGADWY--NLH 244
 DB 254 LSGFSRSRQTDQWVYQSFEEPRRSADPTMLQFKSTPDLRDQGEAAPPGSVDMKATY 313
 QY 245 PLIAHARTGARRVKLHRELAVELEKDLQMRGASVPCGE---LYQ-----WKQ 291
 DB 314 GILREGSSESTSVR--RVSLSVLEK---WQPLVWSSGSTKAVAGQGLTRKVELEQRK 368
 QY 292 LGEDEKDYELVMPWYIWNTRLDKDDNDKVLGMNQELLE-----Y 333
 DB 369 LDESVKQKQLRPSQVGLRQLEKTEEC SRL---QELERKGEAQOQNKELQNMRL 424
 QY 334 FDKYEALR-----ARHSYGPQ----- 349
 DB 425 LDQGEDLRHGLFTQVWELQNKLVHVGPEPAKEVLKLDLETRLELLEVLBEQKQVEBQL 484
 QY 350 --GHRGMSYL---MFESSAT-----GYLEMR--- 371
 DB 485 RLREBELTALKALKEVVASRDQEVHVRQYQYQRTQQLRRSQMDATQDHAVLEAERQRM 544
 QY 372 ----LHRELAEWGLDRIAMGQRSMFSGVRAQLYGFILATQDL----- 410
 DB 545 SALVRGQRLRELTSEETGHW---QSMFOXKEDL---RATQELQLQMEKEMEELG 598
 QY 411 --DIFNQH-----SOGKTR---LKFELSYQEWVVKEL----- 438
 DB 599 EKVTELQRELEQAPASADDTROYAVLKKELLRQGE-ELKTELQERQSEVAGHRHREDE 657
 QY 439 -----RQISENQO---INFRNKLSKQNKAKYLBESLEIMSEK----- 475
 DB 658 KQAVLRYEADRGRLERQNLQKTLQQLRQDCEASKKAWAEATVYIGQRAAVET 717
 QY 476 -LRRFADNRIRYQRTNMQHONREEM-----DAHRRFWDISKQTHERRDAKEENFEM 528
 DB 718 TLRQEBEENDEFRRLILGLEQLKETRGLVNGGAVEARLRDKQLQRLAEKQOLEALNA 777
 QY 529 LQOOE-----RAKY-----VGOQOQINPSSNDCKKAEVSSITFQOE- 568
 DB 778 SQEBSLSLAARALAEARLEAERGLARLGOEQTLNLALEBKQBVLRGRGALEBEQ 837
 QY 569 -----KEMEFYEREMILIKQEKMEKMKRHHHEIFDLEKEP-DEALQQLM 615
 DB 838 KRLLDRTVRLNKELEKIGEDSKQALQQLQQLQLEDKKARREYVADQRAQKWAASEAK 897
 QY 616 YKHGLNEDD 625
 DB 898 TSGGLSRRLD 907
 RESULT 6
 MYH8 HUMAN STANDARD; PRT; 1937 AA.
 AC P13535; Q14910; (Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).
 GN MYH8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90323631; PubMed=2173371;
 RA Karsch-Mizrachi I., Peghail R., Shows T.B. Jr., Leiwand L.A.;
 RT "Generation of a full-length human perinatal myosin heavy-chain-
 RT encoding cDNA.";

```

RL  Gene 89:289-294 (1990).
RN
RP
RP  SEQUENCE FROM N.A.
RX  TISSUE=Skeletal muscle;
RX  MEDLINE=95324556; PubMed=7601129;
RA  Julian E.H., Kelly A.M., Pomidou A.J., Hoffman R., Schiaffino S.,
RA  Steadman H.H., Rubinstein N.A.;
RT  "Characterization of a human perinatal myosin heavy-chain
RT  transcript.";
RL  Eur. J. Biochem. 230:1001-1006(1995).
RN
RN  [3]
RP  SEQUENCE OF 502-1937 FROM N.A.
RC  TISSUE=Skeletal muscle;
RX  MEDLINE=90235862; PubMed=1691980;
RA  Boker E., Buchberger-Seidl A., Braun T., Singh S., Goede H.W.,
RA  Arnold H.H.;
RT  "Identification of three developmentally controlled isoforms of human
RT  myosin heavy chains.";
RL  Eur. J. Biochem. 189:55-65(1990).
RN
RN  [4]
RP  SEQUENCE OF 860-1937 FROM N.A.
RX  MEDLINE=89234168; PubMed=2715179;
RA  Peghail R., Leinwand L.A.;
RT  "Molecular genetic characterization of a developmentally regulated
RT  human perinatal myosin heavy chain.";
RL  J. Cell Biol. 108:1791-1797(1989).
RN
RN  [5]
RP  SEQUENCE OF 1-46 FROM N.A.
RA  Esser K., Tichar A., Myszkowski M.;
RT  "Isolation and characterization of the human perinatal MHC promoter.";
RL  Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
CC
CC  -1- FUNCTION: Muscle contraction.
CC  -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC  heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC  and 2 regulatory light chain subunits (MLC-2).
CC  -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC  -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC  cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC  characteristic for alpha-helical coiled coils.
CC  -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC  meromyosin (LM) and 1 heavy meromyosin (HMM). It can later be
CC  split further into 2 globular subfragments (S1) and 1 rod-shaped
CC  subfragment (S2).
CC  -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC  -1- SIMILARITY: Contains 1 IQ domain.
-----
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
CC
DR  EMBL; M36769; AAC17185.1; -
DR  EMBL; Z38133; CAAB6293.1; -
DR  EMBL; X51592; CA35941.1; -
DR  EMBL; AF067143; AAC21557.1; -
DR  PIR; I38055; I38055.
DR  HSSP; P13538; 2MYS.
DR  Genew; HGNC:7578; MYH8.
DR  MIM; 160741; -
DR  GO; GO:0005859; C:muscle myosin; TAS.
DR  GO; GO:0008307; F:structural constituent of muscle; TAS.
DR  InterPro; IPR000048; IQ_region.
DR  InterPro; IPR001609; myosin head.
DR  InterPro; IPR004009; myosin N.
DR  InterPro; IPR002928; Myosin_tail.
DR  Pfam; PF00612; IQ; 1.
DR  Pfam; PF00063; myosin_head; 1.
DR  Pfam; PF02736; Myosin_N; 1.
DR  Pfam; PF01576; Myosin_tail; 1.
DR  PRINTS; PR00193; MYOSINHEAVY.

```

Query Match	Beet Local	Similarity	5.3%, Score 174; DB 1; Length 1937;	Matches 121; Conservative 19; Pred. No. 0.13; Mismatches 201; Indels 206; Gaps 26
DR	Prodom	PD000355; myosin_head; 1.		
DR	SMART; SM00015; IQ; 1.			
DR	SMART; SM00242; MYSC; 1.			
DR	PROSITE; PS50096; IQ; 1.			
KM	Myosin; muscle protein; Coiled coil; Thick filament; Actin-binding; ATP-binding; Methylation; Multigene family; Calmodulin-binding.			
KM	ATP-binding; Methylation; Multigene family; Calmodulin-binding.			
FT	DOMAIN	1 780		
FT	DOMAIN	781 813		
FT	DOMAIN	842 1937		
FT	NP_BIND	181 188		
FT	DOMAIN	658 680		
FT	DOMAIN	760 774		
FT	MOD_RES	132 132		
FT	CONFLICT	15 15		
FT	CONFLICT	970 970		
FT	CONFLICT	1072 1072		
FT	CONFLICT	1247 1247		
FT	CONFLICT	1251 1252		
FT	CONFLICT	1261 1261		
FT	CONFLICT	1297 1297		
FT	CONFLICT	1377 1378		
FT	CONFLICT	1504 1505		
FT	CONFLICT	1847 1847		
FT	CONFLICT	1914 1914		
FT	SEQUENCE	1937 AA; 222762 MW; A3EB2D151792E9E8 CRC64;		
QY	104 QANISGRALSRKYDNNFVAAPPVSRPPEGGWNNQAGSGAHTAYAQEFPVEDVDN	163		
DB	1280 RARIQTEAGEYSRDLDEKDALVSLSR-----SKQASTQQLIEELKHQLEE	1324		
QY	164 ASEBEND-SDPLDSDS--DDDLASDYPDSVSQKSHGSRKKQKKKKFFGSLDSISIEGIN	220		
DB	1325 ETKKKNALAHLLQSSRHDCDLRLRQYEEBOGKALOR-----ALSKVAN	1368		
QY	221 EPQRMHCPCACQNGPAILWYNLHPLAHARTKARRYKHLRELAEVLEKDLQMGASVI	280		
DB	1369 SEVAQW-----RTKYEITDALQTELEELKAKKGLAQR-----	1399		
QY	281 PCGEIYQWKGGLGDEKDYELVMPWYIIMNT-RLDKDNDKWLGMG-----	326		
DB	1400 -----LQAEHEVEAVNNAKCSLEKTKQRLQNEVEDLMLDVERSNAAACALDKK	1448		
QY	327 ---NOELIEFFDKYELALRAHSHGSPQCHRMSTLMF-----ESSATYGLIEAERL	372		
DB	1449 QRNDDKYLSEKKQYEEBQTEQLEASQKESRSLSTELFKVKNVYESIDQLTELLRRENKNL	1508		
QY	373 HRELAEMGLDRILAWGOKRSMFSGGVRL-----YGFLLTKDLDLIFNGHSGCK	420		
DB	1509 QQELSDL-TEQDLAGGKQ-----THEBEKTKQVBOEKCEIQALAEBAEASLEHBERKI	1561		
QY	421 TRLEFELKSYQEMVAVKELRQISEDNOOLNFKKLSKONKHAUYLESLIEIMSEKLRTA	480		
DB	1562 LRQLDELNAYKSEVDRIKAEKDESIDQL-----KRN-HTRVETMGSITLDAEIR--	1609		
QY	481 EDNRIEIVORTKQHEONREEMDAH-DRFMDSTIK-----QIH-----	516		
DB	1610 SRNDALRVKKMEGDLNEMETQLNHAURLAESIRNRYNTQGLIKETQLHLDALRQED	1669		
QY	517 -----EER-----DAK-EEFPMLOOQORAKVVGQO-----QNTIPSSDD	552		
DB	1670 LKEQALIVERANILLQAEIIEELWATTEQTERSRIKAEQELLDSERVOLLHTQNTSLINT	1729		
QY	553 CRKKAEEVSFFIEFOEKEMEEFVE-----REMLIKQOEKK--MEDM	592		
DB	1730 KKKLENDVSL-----QSEVEEIVIOESNAAEKAKKAITTDAAAMAELKQEDTSANHERM	1785		
QY	593 KGRHHEEIPDLKEKFEALQLMYKHG 619			
DB	1786 KKNLEQTVKDLQHRIDEA-EQLALKGG 1811			

```

RESULT 7
R50. AQUE STANDARD; PRT; 978 AA.
ID R50. AQUE
AC 067124;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AQ_1006.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=981916666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Kellar M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: Belongs to the SMC family. Rad50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000718; AAC07092.1; -.
CC PIR; A70387; A70387.
CC HAMAP; MF 00449; -.
CC DR InterPro:IPR003439; ABC transporter.
CC DR InterPro:IPR007523; SbcC.
CC DR InterPro:IPR004592; SbcC.
CC DR InterPro:IPR002017; Spectrin.
CC DR Pfam; PF04430; DUF498; 1.
CC DR TIGRFAMs; TIGR00618; sbcc; 1.
CC KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC FT NP_BIND 32 39 ATP (By SIMILARITY)
CC FT DOMAIN 160 826 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 978 AA; 115897 MW; 980F2BF51ADD1151 CRC64;
Query Match 5.2%; Score 173.5; DB 1; Length 978;
Best Local Similarity 19.6%; Pred. No. 0.062;
Matches 130; Conservative 107; Mismatches 212; Indels 215; Gaps 30;
QY 106 NISGRGALSRYKD--NNFVAPPVSRPPEGGMNWQARGSAOHTAVOEPFVEDVDVN 163
DB 179 NLRGKRALKEVEYLDYDTP--PTKREYLE-----KTLKNLEELKEIKETEER 225
QY 164 ASEBENDSDALDSDDDLAS-----DDYDSVVS---QKSHSRK----- 139
DB 226 LRDELKKAERKDSLEBELSQVYTKLEINLEKEVEKLEKLEFSRYAVPYPIAKIEE 285
QY 200 -----QNKWFKKFPFGSLDLSI--EQIN-----EQRQWHPCAQNGP 235
DB 286 IDKKLTLEKRYKNTKTELAVALDELSFAGEELNRIRAEKEKPEKEKEKEKE----- 337
QY 236 GAIDWYNLHPLLAARTGARRVYLHLELAEVLEKDIQMGASVYPCGEIYGQWKGIGED 295

```

```

DB 338 -----LEH-----RLKLGQEIKEIL-KELSQLSSS-----LKEK 365
QY 236 EKDYETVMPWYIIMNTRLDKDNDKWLGMGNQELLEYF-----DKYEALRARSYGPQ 349
DB 366 EREYFOAQOEPB-DLSERVEK---OKLVATEERLEKIKELFSSEETLSLKMERRLVE 421
QY 350 GHRGNSVL-----MESSATGYLEAERLH-----RELAMGLDRJAWGQKSMF 393
DB 422 LQRLKELEKEGQLENLTQYKEKKVHEKVLNELKELEBELKREIRLHYHNVASYLS 481
QY 394 SGGVNQLYGFATKQDLDIFPHQSGCKTRKE--ELKSYQM----- 433
DB 482 PDTCPCVCGGIYRGALB--NVDAGISLELHNAKEKEKEBEIDTTLKLYAQKINSIKE 539
QY 434 -----VYKELQISEDN-----QQLNYFKKLS-KONKAKVLEB----- 467
DB 540 EMEKLRNVEBELRKIPENLKERIKLEBELREKELHKLNKRYKALJEDRQKQKEAQA 599
QY 468 -----SLEIWSKLRRTAEDNRIVRQTYKQH---EQNREMDAHRFMDSIKQIH 516
DB 600 KLHKAQTELELKEKLR---EKSLVKEFKELRYVERLEDEYESLKEEINYNSYLOEIE 656
QY 517 ERDRAKEEFEMLQOOERAKVVGQOQONINPSNDCKRAEVSFTIEQE--KEMEEL 574
DB 657 EKEKLRKHFELSSR-KSKLEGELSA-LNESINSLEERKEKLELANIYVAKSPBV 714
QY 575 VE-----EEMLIKQOKEKMEDMKKRHH-----EIFPLEKEFDALBQL 614
DB 715 VELYIGDKAELEKRIKEFEBSFQSLKKSIEKLEKYEIGRIELSDIKGYESVKQL 774
QY 615 MYKH 618
DB 775 EEKH 778
RESULT 8
MYSS_RABIT STANDARD; PRT; 1084 AA.
ID MYSS_RABIT
AC P02562;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle (fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE OF 1-258.
RA Capony J.-P., Elzinga M.;
RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of
RT myosin."
RL Biophys. J. 33:148A-148A(1981).
RN [2]
RP SEQUENCE OF 259-428.
RX MEDLINE=85131142; PubMed=3972832;
RA Lu R.C., Wong A.;
RT "The amino acid sequence and stability predictions of the hinge
RT region in myosin subfragment 2."
RL J. Biol. Chem. 260:3456-3461(1985).
RN [3]
RP SEQUENCE OF 409-1084 FROM N.A.
RX MEDLINE=87304245; PubMed=3305014;
RA Meda K., Szakiel G., Wittmhofer A.;
RT "Characterization of cDNA coding for the complete light meromyosin
RT portion of a rabbit fast skeletal muscle myosin heavy chain."
RL Eur. J. Biochem. 167:97-102(1987).
CC -1- FUNCTION: Muscle contraction.
CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing

```

cycles of a 28-residue repeat pattern composed of 4 heptapeptides, characteristic for alpha-helical coiled coils.
 -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be split further into 2 globular subfragments (S1) and 1 rod-shaped subfragment (S2).
 -1- SIMILARITY: Contains 1 myosin-like globular head domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL, on the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X05958; CAA29391.1; --
 DR PIR: A02985; A02985.
 DR PIR: A05280; A05280.
 DR PIR: S00084; S00084.
 KM Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Multigene family.
 FT NON_TER 1
 FT DOMAIN 1
 FT NON_CONS 258 259 ALPHA-HELICAL TAILPIECE (S2).
 FT DOMAIN 258 259 RODLIKE TAIL (S2 AND LMW DOMAINS).
 FT DOMAIN 455 1084 COILED COIL (POTENTIAL).
 FT VARIANT 405 405 L -> V.
 FT VARIANT 408 408 V -> L.
 FT VARIANT 421 421 E -> D.
 FT VARIANT 423 423 S -> G.
 FT VARIANT 426 426 K -> R.
 SQ SEQUENCE 1084 AA; 12548 MW; 229CFD69A6B1F7F0 CRC64;

 Query Match 5.2%; Score 173; DB 1; Length 1084;
 Best Local Similarity 21.1%; Pred. No. 0.074;
 Matches 124; Conservative 91; Mismatches 184; Indels 188; Gaps 30;
 141 ARGGSNQHTAVQEPFVEDVDVDAASEND-----SDALDDSD--DDLDASDDVDVSQK 193
 451 SRGQ-----QAFQOIEGLKQLEBETKAKSALHALLQSSRRDCDLREYEEQEAQ 503
 194 SHSGRKQNTKFKFEGSLDSLSTEQINEPQRMHCPACONGPAILDWYNLHPLAARTK 253
 504 AELQGRMSK-----ANSEVSQWR-TKCEI-----DAIQTEELERAKK 541
 254 GARRVKHLBELALEKDLQMGASVYPCGEIYQWKGJGDEKDEYIWPVVIIMT- 312
 542 LAQRLQDAEEHVAVNS-----KCAELEKTKQRLQNEADL-----MIDVERSN 585
 313 -----RLDKDNDKWLGMGQELLEFPDYALAAHRSYSGQGRGMSVLMFE-----SSA 363
 586 ATCARDKKQKN-----FDKVLAEWKHKHETAELEASQKRSRSLTEFKYKMAVEES 640
 364 TGYLEA-----ERLHELAEMLGDIRIANG-----OKRSMFSGGVRLQYGF 403
 641 LDHLETLKREKNKLOGEISL-TEQIALESAKTHIELEKVKQKQIDQKSELQALAEKES 699
 404 LATKQDLDIFNHSQSK-TRLEKTELSYQDVVVKELRQISSENOQNLNFKNLSKONKHA 462
 700 L-----EHEEKILRIQLLELNQVSEIRKIAEKKEIDQL-----KEN-HL 740
 463 KYLESELIMSEKLTARTEADNRIVRQTKQHQHONNEEMAH--DEFEMDSIK----- 513
 741 RVVESNOSTDAEIR--SRNDLRIRKKQKQEGDNLNEMETQLNANNOAMALNRLNTQG 797
 514 -----QIH-----ERR-----DAK-EENFEMILQOQERAKVGGQOQ----- 542
 798 ILKDQTLHDADVAGDDHKEQJLAWERRRANLMOALIEELRASLEQTERRRRAADDL 857
 543 -----QNPSNDDCKRAEVSSEFIEQEKEMEVEE----- 577
 858 ASERVQLHTQN---TSLINTKKKLETDISQI---QGEEMEDIVQEARNAEKAATDA 911

578 ---REMLIKDQEKK--MEDMKRHHHEFDLEKFPDEALQOLMYKKG 619
 912 ANMAEELKKEQDTSALERKMKMEQTVADLQQRLEDA-EQLAKKG 957
 Db
 RESULT 9
 TARA HUMAN
 ID TARA HUMAN STANDARD; PRT; 593 AA.
 AC Q9H2B6; Q94797; Q96DM1; Q9BTL7; Q9BTL7; Q9B938; Q9Y1L4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE TF10 and F-actin binding protein (Protein Tara) (Trio-associated repeat on actin) (HRIHF8122).
 GN TRIOBP OR TARA OR KIAA1662.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21067080; PubMed=11148140;
 RA Seipel K., O'Brien S.P., Iannotti E., Medley Q.G., Streuli M.,
 RT "Tara, a novel F-actin binding protein, associates with the Trio
 RT quantin nucleotide exchange factor and regulates actin cytoskeletal
 RT organization".
 RL J. Cell Sci. 114:389-399(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones".
 RL DNA Res. 9:99-106(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=21156230; PubMed=11258795;
 RA Hiroawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
 RT "Identification of novel transcribed sequences on human chromosome 22
 RT by expressed sequence tag mapping".
 RL DNA Res. 8:1-9(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clamp M., Smith L.J., Alnscough R., Almeida J.P., Babbage A.K.,
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gillbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Senter H.K., Skuce C.D., Smalley S., Smith W.L.,
 RA Soudelund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshitaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.T.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,

RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutou K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RT discoidium complexed with MgADP.Befx and MgADP.Alfa-";
RL Biochemistry 34:8960-8972 (1995).
RN [5]
RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RP MEDLINE=95345067; PubMed=7619796;
RX Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT truncated head of Dictyostelium discoidium myosin to 2.7-A
RT resolution.";
RL Biochemistry 34:8973-8981 (1995).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-ADP.vanadate complex of the
RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417 (1996).
RN [7]
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RP MEDLINE=97452580; PubMed=9305951;
RX Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammass, and MgAMPNP complexes
RT of the Dictyostelium discoidium myosin motor domain.";
RL Biochemistry 36:11619-11628 (1997).
RN [8]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bageshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
RT of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT Dictyostelium discoidium myosin motor domain.";
RL J. Mol. Biol. 274:394-407 (1997).
RN [9]
RN FUNCTION: Myosin is a protein that binds to actin and has ATPase
RT activity that is activated by actin.
RT SUBUNIT: Myosin II heavy chain is two-headed. It self-assembles
RT into filaments. Hexamer of 2 heavy chain subunits (MHC), 2 alkali
RT light chain subunits (MLC) and 2 regulatory light chain subunits
RT (MLC-2).
RN [10]
RN SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
RN CORTEX.
RN [11]
RN DOMAIN: Each myosin heavy chain can be split into 1 light
RN meromyosin (LMM) and 1 heavy meromyosin (HMM). It can be further
RN split into 2 globular subfragments (S1) and 1 rod-shaped
RN subfragment (S2).
RN [12]
RN DOMAIN: The rodlike tail sequence is highly repetitive, showing
RN cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
RN characteristic for alpha-helical coiled coils.
RN [13]
RN PTM: Phosphorylation inhibits thick filament formation and reduces
RN the actin-activated ATPase activity.
RN [14]
RN MISCELLANEOUS: Dictyostelium myosin II has no K(2)EDTA ATPase
RN activity, perhaps correlated with the absence of a Cys at the SH-1
RN position (688).
RN [15]
RN SIMILARITY: Contains 1 myosin-like globular head domain.
RN [16]
RN SIMILARITY: Contains 1 IQ domain.
RN [17]
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN [18]
RN EMBL: M14628; AAA3227.1; -
RN PIR: A26655; A26655.
RN PDB: 1MMA; 03-DEC-97.
RN PDB: 1MMD; 17-AUG-96.
RN PDB: 1MNG; 03-DEC-97.
RN PDB: 1MNN; 03-DEC-97.
RN PDB: 1MND; 17-AUG-96.
RN [19]
RN PDB; 1MNE; 17-AUG-96.
RN PDB; 1VOM; 23-DEC-96.
RN PDB; 1IVK; 28-JAN-98.
RN PDB; 1DOX; 20-DEC-00.
RN PDB; 1DOY; 20-DEC-00.
RN PDB; 1DOZ; 20-DEC-00.
RN PDB; 1DIA; 20-DEC-00.
RN PDB; 1DIB; 20-DEC-00.
RN PDB; 1DIC; 20-DEC-00.
RN PDB; 1FMV; 20-DEC-00.
RN PDB; 1FMW; 20-DEC-00.
RN PDB; 1G8X; 17-JAN-01.
RN PDB; 1UWY; 07-NOV-01.
RN PDB; 1UX2; 07-NOV-01.
RN DictyBase; DDB0002015; mbca.
RN DictyBase; DDB0000048; IQ_region.
RN InterPro; IPR001609; myosin_head.
RN InterPro; IPR004009; Myosin_N.
RN InterPro; IPR008989; Myosin_S1_N.
RN Pfam; PF00612; IQ; 2.
RN Pfam; PF00063; myosin_head; 1.
RN Pfam; PF02736; Myosin_N; 1.
RN PRINTS; PR00193; MYOSTINHEAVY.
RN PRODOM; PD000355; myosin_head; 1.
RN SMART; SM00015; IQ; 1.
RN SMART; SM00242; MYSC; 1.
RN PROSITE; PS50096; IQ; 1.
RN Myosin; Coiled coil; Actin-binding; 3D-structure;
RN Calmodulin-binding; Methylation; Phosphorylation.
RN MYOSIN HEAD-LIKE.
RN [20]
RN COILED COIL (POTENTIAL).
RN [21]
RN ATP.
RN [22]
RN ACTIN-BINDING.
RN [23]
RN ACTIN-BINDING.
RN [24]
RN METHYLATION (DI-) (POTENTIAL).
RN [25]
RN PHOSPHORYLATION (BY MHCK).
RN [26]
RN PHOSPHORYLATION (BY MHCK).
RN [27]
RN PHOSPHORYLATION (BY MHCK).
RN [28]
RN PHOSPHORYLATION (BY MHCK).
RN [29]
RN PHOSPHORYLATION (BY MHCK).
RN [30]
RN PHOSPHORYLATION (BY MHCK).
RN [31]
RN PHOSPHORYLATION (BY MHCK).
RN [32]
RN PHOSPHORYLATION (BY MHCK).
RN [33]
RN PHOSPHORYLATION (BY MHCK).
RN [34]
RN PHOSPHORYLATION (BY MHCK).
RN [35]
RN PHOSPHORYLATION (BY MHCK).
RN [36]
RN PHOSPHORYLATION (BY MHCK).
RN [37]
RN PHOSPHORYLATION (BY MHCK).
RN [38]
RN PHOSPHORYLATION (BY MHCK).
RN [39]
RN PHOSPHORYLATION (BY MHCK).
RN [40]
RN PHOSPHORYLATION (BY MHCK).
RN [41]
RN PHOSPHORYLATION (BY MHCK).
RN [42]
RN PHOSPHORYLATION (BY MHCK).
RN [43]
RN PHOSPHORYLATION (BY MHCK).
RN [44]
RN PHOSPHORYLATION (BY MHCK).
RN [45]
RN PHOSPHORYLATION (BY MHCK).
RN [46]
RN PHOSPHORYLATION (BY MHCK).
RN [47]
RN PHOSPHORYLATION (BY MHCK).
RN [48]
RN PHOSPHORYLATION (BY MHCK).
RN [49]
RN PHOSPHORYLATION (BY MHCK).
RN [50]
RN PHOSPHORYLATION (BY MHCK).
RN [51]
RN PHOSPHORYLATION (BY MHCK).
RN [52]
RN PHOSPHORYLATION (BY MHCK).
RN [53]
RN PHOSPHORYLATION (BY MHCK).
RN [54]
RN PHOSPHORYLATION (BY MHCK).
RN [55]
RN PHOSPHORYLATION (BY MHCK).
RN [56]
RN PHOSPHORYLATION (BY MHCK).
RN [57]
RN PHOSPHORYLATION (BY MHCK).
RN [58]
RN PHOSPHORYLATION (BY MHCK).
RN [59]
RN PHOSPHORYLATION (BY MHCK).
RN [60]
RN PHOSPHORYLATION (BY MHCK).
RN [61]
RN PHOSPHORYLATION (BY MHCK).
RN [62]
RN PHOSPHORYLATION (BY MHCK).
RN [63]
RN PHOSPHORYLATION (BY MHCK).
RN [64]
RN PHOSPHORYLATION (BY MHCK).
RN [65]
RN PHOSPHORYLATION (BY MHCK).
RN [66]
RN PHOSPHORYLATION (BY MHCK).
RN [67]
RN PHOSPHORYLATION (BY MHCK).
RN [68]
RN PHOSPHORYLATION (BY MHCK).
RN [69]
RN PHOSPHORYLATION (BY MHCK).
RN [70]
RN PHOSPHORYLATION (BY MHCK).
RN [71]
RN PHOSPHORYLATION (BY MHCK).
RN [72]
RN PHOSPHORYLATION (BY MHCK).
RN [73]
RN PHOSPHORYLATION (BY MHCK).
RN [74]
RN PHOSPHORYLATION (BY MHCK).
RN [75]
RN PHOSPHORYLATION (BY MHCK).
RN [76]
RN PHOSPHORYLATION (BY MHCK).
RN [77]
RN PHOSPHORYLATION (BY MHCK).
RN [78]
RN PHOSPHORYLATION (BY MHCK).
RN [79]
RN PHOSPHORYLATION (BY MHCK).
RN [80]
RN PHOSPHORYLATION (BY MHCK).
RN [81]
RN PHOSPHORYLATION (BY MHCK).
RN [82]
RN PHOSPHORYLATION (BY MHCK).
RN [83]
RN PHOSPHORYLATION (BY MHCK).
RN [84]
RN PHOSPHORYLATION (BY MHCK).
RN [85]
RN PHOSPHORYLATION (BY MHCK).
RN [86]
RN PHOSPHORYLATION (BY MHCK).
RN [87]
RN PHOSPHORYLATION (BY MHCK).
RN [88]
RN PHOSPHORYLATION (BY MHCK).
RN [89]
RN PHOSPHORYLATION (BY MHCK).
RN [90]
RN PHOSPHORYLATION (BY MHCK).
RN [91]
RN PHOSPHORYLATION (BY MHCK).
RN [92]
RN PHOSPHORYLATION (BY MHCK).
RN [93]
RN PHOSPHORYLATION (BY MHCK).
RN [94]
RN PHOSPHORYLATION (BY MHCK).
RN [95]
RN PHOSPHORYLATION (BY MHCK).
RN [96]
RN PHOSPHORYLATION (BY MHCK).
RN [97]
RN PHOSPHORYLATION (BY MHCK).
RN [98]
RN PHOSPHORYLATION (BY MHCK).
RN [99]
RN PHOSPHORYLATION (BY MHCK).
RN [100]
RN PHOSPHORYLATION (BY MHCK).
RN [101]
RN PHOSPHORYLATION (BY MHCK).
RN [102]
RN PHOSPHORYLATION (BY MHCK).
RN [103]
RN PHOSPHORYLATION (BY MHCK).
RN [104]
RN PHOSPHORYLATION (BY MHCK).
RN [105]
RN PHOSPHORYLATION (BY MHCK).
RN [106]
RN PHOSPHORYLATION (BY MHCK).
RN [107]
RN PHOSPHORYLATION (BY MHCK).
RN [108]
RN PHOSPHORYLATION (BY MHCK).
RN [109]
RN PHOSPHORYLATION (BY MHCK).
RN [110]
RN PHOSPHORYLATION (BY MHCK).
RN [111]
RN PHOSPHORYLATION (BY MHCK).
RN [112]
RN PHOSPHORYLATION (BY MHCK).
RN [113]
RN PHOSPHORYLATION (BY MHCK).
RN [114]
RN PHOSPHORYLATION (BY MHCK).
RN [115]
RN PHOSPHORYLATION (BY MHCK).
RN [116]
RN PHOSPHORYLATION (BY MHCK).
RN [117]
RN PHOSPHORYLATION (BY MHCK).
RN [118]
RN PHOSPHORYLATION (BY MHCK).
RN [119]
RN PHOSPHORYLATION (BY MHCK).
RN [120]
RN PHOSPHORYLATION (BY MHCK).
RN [121]
RN PHOSPHORYLATION (BY MHCK).
RN [122]
RN PHOSPHORYLATION (BY MHCK).
RN [123]
RN PHOSPHORYLATION (BY MHCK).
RN [124]
RN PHOSPHORYLATION (BY MHCK).
RN [125]
RN PHOSPHORYLATION (BY MHCK).
RN [126]
RN PHOSPHORYLATION (BY MHCK).
RN [127]
RN PHOSPHORYLATION (BY MHCK).
RN [128]
RN PHOSPHORYLATION (BY MHCK).
RN [129]
RN PHOSPHORYLATION (BY MHCK).
RN [130]
RN PHOSPHORYLATION (BY MHCK).
RN [131]
RN PHOSPHORYLATION (BY MHCK).
RN [132]
RN PHOSPHORYLATION (BY MHCK).
RN [133]
RN PHOSPHORYLATION (BY MHCK).
RN [134]
RN PHOSPHORYLATION (BY MHCK).
RN [135]
RN PHOSPHORYLATION (BY MHCK).
RN [136]
RN PHOSPHORYLATION (BY MHCK).
RN [137]
RN PHOSPHORYLATION (BY MHCK).
RN [138]
RN PHOSPHORYLATION (BY MHCK).
RN [139]
RN PHOSPHORYLATION (BY MHCK).
RN [140]
RN PHOSPHORYLATION (BY MHCK).
RN [141]
RN PHOSPHORYLATION (BY MHCK).
RN [142]
RN PHOSPHORYLATION (BY MHCK).
RN [143]
RN PHOSPHORYLATION (BY MHCK).
RN [144]
RN PHOSPHORYLATION (BY MHCK).
RN [145]
RN PHOSPHORYLATION (BY MHCK).
RN [146]
RN PHOSPHORYLATION (BY MHCK).
RN [147]
RN PHOSPHORYLATION (BY MHCK).
RN [148]
RN PHOSPHORYLATION (BY MHCK).
RN [149]
RN PHOSPHORYLATION (BY MHCK).
RN [150]
RN PHOSPHORYLATION (BY MHCK).
RN [151]
RN PHOSPHORYLATION (BY MHCK).
RN [152]
RN PHOSPHORYLATION (BY MHCK).
RN [153]
RN PHOSPHORYLATION (BY MHCK).
RN [154]
RN PHOSPHORYLATION (BY MHCK).
RN [155]
RN PHOSPHORYLATION (BY MHCK).
RN [156]
RN PHOSPHORYLATION (BY MHCK).
RN [157]
RN PHOSPHORYLATION (BY MHCK).
RN [158]
RN PHOSPHORYLATION (BY MHCK).
RN [159]
RN PHOSPHORYLATION (BY MHCK).
RN [160]
RN PHOSPHORYLATION (BY MHCK).
RN [161]
RN PHOSPHORYLATION (BY MHCK).
RN [162]
RN PHOSPHORYLATION (BY MHCK).
RN [163]
RN PHOSPHORYLATION (BY MHCK).
RN [164]
RN PHOSPHORYLATION (BY MHCK).
RN [165]
RN PHOSPHORYLATION (BY MHCK).
RN [166]
RN PHOSPHORYLATION (BY MHCK).
RN [167]
RN PHOSPHORYLATION (BY MHCK).
RN [168]
RN PHOSPHORYLATION (BY MHCK).
RN [169]
RN PHOSPHORYLATION (BY MHCK).
RN [170]
RN PHOSPHORYLATION (BY MHCK).
RN [171]
RN PHOSPHORYLATION (BY MHCK).
RN [172]
RN PHOSPHORYLATION (BY MHCK).
RN [173]
RN PHOSPHORYLATION (BY MHCK).
RN [174]
RN PHOSPHORYLATION (BY MHCK).
RN [175]
RN PHOSPHORYLATION (BY MHCK).
RN [176]
RN PHOSPHORYLATION (BY MHCK).
RN [177]
RN PHOSPHORYLATION (BY MHCK).
RN [178]
RN PHOSPHORYLATION (BY MHCK).
RN [179]
RN PHOSPHORYLATION (BY MHCK).
RN [180]
RN PHOSPHORYLATION (BY MHCK).
RN [181]
RN PHOSPHORYLATION (BY MHCK).
RN [182]
RN PHOSPHORYLATION (BY MHCK).
RN [183]
RN PHOSPHORYLATION (BY MHCK).
RN [184]
RN PHOSPHORYLATION (BY MHCK).
RN [185]
RN PHOSPHORYLATION (BY MHCK).
RN [186]
RN PHOSPHORYLATION (BY MHCK).
RN [187]
RN PHOSPHORYLATION (BY MHCK).
RN [188]
RN PHOSPHORYLATION (BY MHCK).
RN [189]
RN PHOSPHORYLATION (BY MHCK).
RN [190]
RN PHOSPHORYLATION (BY MHCK).
RN [191]
RN PHOSPHORYLATION (BY MHCK).
RN [192]
RN PHOSPHORYLATION (BY MHCK).
RN [193]
RN PHOSPHORYLATION (BY MHCK).
RN [194]
RN PHOSPHORYLATION (BY MHCK).
RN [195]
RN PHOSPHORYLATION (BY MHCK).
RN [196]
RN PHOSPHORYLATION (BY MHCK).
RN [197]
RN PHOSPHORYLATION (BY MHCK).
RN [198]
RN PHOSPHORYLATION (BY MHCK).
RN [199]
RN PHOSPHORYLATION (BY MHCK).
RN [200]
RN PHOSPHORYLATION (BY MHCK).
RN [201]
RN PHOSPHORYLATION (BY MHCK).
RN [202]
RN PHOSPHORYLATION (BY MHCK).
RN [203]
RN PHOSPHORYLATION (BY MHCK).
RN [204]
RN PHOSPHORYLATION (BY MHCK).
RN [205]
RN PHOSPHORYLATION (BY MHCK).
RN [206]
RN PHOSPHORYLATION (BY MHCK).
RN [207]
RN PHOSPHORYLATION (BY MHCK).
RN [208]
RN PHOSPHORYLATION (BY MHCK).
RN [209]
RN PHOSPHORYLATION (BY MHCK).
RN [210]
RN PHOSPHORYLATION (BY MHCK).
RN [211]
RN PHOSPHORYLATION (BY MHCK).
RN [212]
RN PHOSPHORYLATION (BY MHCK).
RN [213]
RN PHOSPHORYLATION (BY MHCK).
RN [214]
RN PHOSPHORYLATION (BY MHCK).
RN [215]
RN PHOSPHORYLATION (BY MHCK).
RN [216]
RN PHOSPHORYLATION (BY MHCK).
RN [217]
RN PHOSPHORYLATION (BY MHCK).
RN [218]
RN PHOSPHORYLATION (BY MHCK).
RN [219]
RN PHOSPHORYLATION (BY MHCK).
RN [220]
RN PHOSPHORYLATION (BY MHCK).
RN [221]
RN PHOSPHORYLATION (BY MHCK).
RN [222]
RN PHOSPHORYLATION (BY MHCK).
RN [223]
RN PHOSPHORYLATION (BY MHCK).
RN [224]
RN PHOSPHORYLATION (BY MHCK).
RN [225]
RN PHOSPHORYLATION (BY MHCK).
RN [226]
RN PHOSPHORYLATION (BY MHCK).
RN [227]
RN PHOSPHORYLATION (BY MHCK).
RN [228]
RN PHOSPHORYLATION (BY MHCK).
RN [229]
RN PHOSPHORYLATION (BY MHCK).
RN [230]
RN PHOSPHORYLATION (BY MHCK).
RN [231]
RN PHOSPHORYLATION (BY MHCK).
RN [232]
RN PHOSPHORYLATION (BY MHCK).
RN [233]
RN PHOSPHORYLATION (BY MHCK).
RN [234]
RN PHOSPHORYLATION (BY MHCK).
RN [235]
RN PHOSPHORYLATION (BY MHCK).
RN [236]
RN PHOSPHORYLATION (BY MHCK).
RN [237]
RN PHOSPHORYLATION (BY MHCK).
RN [238]
RN PHOSPHORYLATION (BY MHCK).
RN [239]
RN PHOSPHORYLATION (BY MHCK).
RN [240]
RN PHOSPHORYLATION (BY MHCK).
RN [241]
RN PHOSPHORYLATION (BY MHCK).
RN [242]
RN PHOSPHORYLATION (BY MHCK).
RN [243]
RN PHOSPHORYLATION (BY MHCK).
RN [244]
RN PHOSPHORYLATION (BY MHCK).
RN [245]
RN PHOSPHORYLATION (BY MHCK).
RN [246]
RN PHOSPHORYLATION (BY MHCK).
RN [247]
RN PHOSPHORYLATION (BY MHCK).
RN [248]
RN PHOSPHORYLATION (BY MHCK).
RN [249]
RN PHOSPHORYLATION (BY MHCK).
RN [250]
RN PHOSPHORYLATION (BY MHCK).
RN [251]
RN PHOSPHORYLATION (BY MHCK).
RN [252]
RN PHOSPHORYLATION (BY MHCK).
RN [253]
RN PHOSPHORYLATION (BY MHCK).
RN [254]
RN PHOSPHORYLATION (BY MHCK).
RN [255]
RN PHOSPHORYLATION (BY MHCK).
RN [256]
RN PHOSPHORYLATION (BY MHCK).
RN [257]
RN PHOSPHORYLATION (BY MHCK).
RN [258]
RN PHOSPHORYLATION (BY MHCK).
RN [259]
RN PHOSPHORYLATION (BY MHCK).
RN [260]
RN PHOSPHORYLATION (BY MHCK).
RN [261]
RN PHOSPHORYLATION (BY MHCK).
RN [262]
RN PHOSPHORYLATION (BY MHCK).
RN [263]
RN PHOSPHORYLATION (BY MHCK).
RN [264]
RN PHOSPHORYLATION (BY MHCK).
RN [265]
RN PHOSPHORYLATION (BY MHCK).
RN [266]
RN PHOSPHORYLATION (BY MHCK).
RN [267]
RN PHOSPHORYLATION (BY MHCK).
RN [268]
RN PHOSPHORYLATION (BY MHCK).
RN [269]
RN PHOSPHORYLATION (BY MHCK).
RN [270]
RN PHOSPHORYLATION (BY MHCK).
RN [271]
RN PHOSPHORYLATION (BY MHCK).
RN [272]
RN PHOSPHORYLATION (BY MHCK).
RN [273]
RN PHOSPHORYLATION (BY MHCK).
RN [274]
RN PHOSPHORYLATION (BY MHCK).
RN [275]
RN PHOSPHORYLATION (BY MHCK).
RN [276]
RN PHOSPHORYLATION (BY MHCK).
RN [277]
RN PHOSPHORYLATION (BY MHCK).
RN [278]
RN PHOSPHORYLATION (BY MHCK).
RN [279]
RN PHOSPHORYLATION (BY MHCK).
RN [280]
RN PHOSPHORYLATION (BY MHCK).
RN [281]
RN PHOSPHORYLATION (BY MHCK).
RN [282]
RN PHOSPHORYLATION (BY MHCK).
RN [283]
RN PHOSPHORYLATION (BY MHCK).
RN [284]
RN PHOSPHORYLATION (BY MHCK).
RN [285]
RN PHOSPHORYLATION (BY MHCK).
RN [286]
RN PHOSPHORYLATION (BY MHCK).
RN [287]
RN PHOSPHORYLATION (BY MHCK).
RN [288]
RN PHOSPHORYLATION (BY MHCK).
RN [289]
RN PHOSPHORYLATION (BY MHCK).
RN [290]
RN PHOSPHORYLATION (BY MHCK).
RN [291]
RN PHOSPHORYLATION (BY MHCK).
RN [292]
RN PHOSPHORYLATION (BY MHCK).
RN [293]
RN PHOSPHORYLATION (BY MHCK).
RN [294]
RN PHOSPHORYLATION (BY MHCK).
RN [295]
RN PHOSPHORYLATION (BY MHCK).
RN [296]
RN PHOSPHORYLATION (BY MHCK).
RN [297]
RN PHOSPHORYLATION (BY MHCK).
RN [298]
RN PHOSPHORYLATION (BY MHCK).
RN [299]
RN PHOSPHORYLATION (BY MHCK).
RN [300]
RN PHOSPHORYLATION (BY MHCK).
RN [301]
RN PHOSPHORYLATION (BY MHCK).
RN [302]
RN PHOSPHORYLATION (BY MHCK).
RN [303]
RN PHOSPHORYLATION (BY MHCK).
RN [304]
RN PHOSPHORYLATION (BY MHCK).
RN [305]
RN PHOSPHORYLATION (BY MHCK).
RN [306]
RN PHOSPHORYLATION (BY MHCK).
RN [307]
RN PHOSPHORYLATION (BY MHCK).
RN [308]
RN PHOSPHORYLATION (BY MHCK).
RN [309]
RN PHOSPHORYLATION (BY MHCK).
RN [310]
RN PHOSPHORYLATION (BY MHCK).
RN [311]
RN PHOSPHORYLATION (BY MHCK).
RN [312]
RN PHOSPHORYLATION (BY MHCK).
RN [313]
RN PHOSPHORYLATION (BY MHCK).
RN [314]
RN PHOSPHORYLATION (BY MHCK).
RN [315]
RN PHOSPHORYLATION (BY MHCK).
RN [316]
RN PHOSPHORYLATION (BY MHCK).
RN [317]
RN PHOSPHORYLATION (BY MHCK).
RN [318]
RN PHOSPHORYLATION (BY MHCK).
RN [319]
RN PHOSPHORYLATION (BY MHCK).
RN [320]
RN PHOSPHORYLATION (BY MHCK).
RN [321]
RN PHOSPHORYLATION (BY MHCK).
RN [322]
RN PHOSPHORYLATION (BY MHCK).
RN [323]
RN PHOSPHORYLATION (BY MHCK).
RN [324]
RN PHOSPHORYLATION (BY MHCK).
RN [325]
RN PHOSPHORYLATION (BY MHCK).
RN [326]
RN PHOSPHORYLATION (BY MHCK).
RN [327]
RN PHOSPHORYLATION (BY MHCK).
RN [328]
RN PHOSPHORYLATION (BY MHCK).
RN [329]
RN PHOSPHORYLATION (BY MHCK).
RN [330]
RN PHOSPHORYLATION (BY MHCK).
RN [331]
RN PHOSPHORYLATION (BY MHCK).
RN [332]
RN PHOSPHORYLATION (BY MHCK).
RN [333]
RN PHOSPHORYLATION (BY MHCK).
RN [334]
RN PHOSPHORYLATION (BY MHCK).
RN [335]
RN PHOSPHORYLATION (BY MHCK).
RN [336]
RN PHOSPHORYLATION (BY MHCK).
RN [337]
RN PHOSPHORYLATION (BY MHCK).
RN [338]
RN PHOSPHORYLATION (BY MHCK).
RN [339]
RN PHOSPHORYLATION (BY MHCK).
RN [340]
RN PHOSPHORYLATION (BY MHCK).
RN [341]
RN PHOSPHORYLATION (BY MHCK).
RN [342]
RN PHOSPHORYLATION (BY MHCK).
RN [343]
RN PHOSPHORYLATION (BY MHCK).
RN [344]
RN PHOSPHORYLATION (BY MHCK).
RN [345]
RN PHOSPHORYLATION (BY MHCK).
RN [346]
RN PHOSPHORYLATION (BY MHCK).
RN [347]
RN PHOSPHORYLATION (BY MHCK).
RN [348]
RN PHOSPHORYLATION (BY MHCK).
RN [349]
RN PHOSPHORYLATION (BY MHCK).
RN [350]
RN PHOSPHORYLATION (BY MHCK).
RN [351]
RN PHOSPHORYLATION (BY MHCK).
RN [352]
RN PHOSPHORYLATION (BY MHCK).
RN [353]
RN PHOSPHORYLATION (BY MHCK).
RN [354]
RN PHOSPHORYLATION (BY MHCK).
RN [355]
RN PHOSPHORYLATION (BY MHCK).
RN [356]
RN PHOSPHORYLATION (BY MHCK).
RN [357]
RN PHOSPHORYLATION (BY MHCK).
RN [358]
RN PHOSPHORYLATION (BY MHCK).
RN [359]
RN PHOSPHORYLATION (BY MHCK).
RN [360]
RN PHOSPHORYLATION (BY MHCK).
RN [361]
RN PHOSPHORYLATION (BY MHCK).
RN [362]
RN PHOSPHORYLATION (BY MHCK).
RN [363]
RN PHOSPHORYLATION (BY MHCK).
RN [364]
RN PHOSPHORYLATION (BY MHCK).
RN [365]
RN PHOSPHORYLATION (BY MHCK).
RN [366]
RN PHOSPHORYLATION (BY MHCK).
RN [367]
RN PHOSPHORYLATION (BY MHCK).
RN [368]
RN PHOSPHORYLATION (BY MHCK).
RN [369]
RN PHOSPHORYLATION (BY MHCK).
RN [370]
RN PHOSPHORYLATION (BY MHCK).
RN [371]
RN PHOSPHORYLATION (BY MHCK).
RN [372]
RN PHOSPHORYLATION (BY MHCK).
RN [373]
RN PHOSPHORYLATION (BY MHCK).
RN [374]
RN PHOSPHORYLATION (BY MHCK).
RN [375]
RN PHOSPHORYLATION (BY MHCK).
RN [376]
RN PHOSPHORYLATION (BY MHCK).
RN [377]
RN PHOSPHORYLATION (BY MHCK).
RN [378]
RN PHOSPHORYLATION (BY MHCK).
RN [379]
RN PHOSPHORYLATION (BY MHCK).
RN [380]
RN PHOSPHORYLATION (BY MHCK).
RN [381]
RN PHOSPHORYLATION (BY MHCK).
RN [382]
RN PHOSPHORYLATION (BY MHCK).
RN [383]
RN PHOSPHORYLATION (BY MHCK).
RN [384]
RN PHOSPHORYLATION (BY MHCK).
RN [385]
RN PHOSPHORYLATION (BY MHCK).
RN [386]
RN PHOSPHORYLATION (BY MHCK).
RN [387]
RN PHOSPHORYLATION (BY MHCK).
RN [388]
RN PHOSPHORYLATION (BY MHCK).
RN [389]
RN PHOSPHORYLATION (BY MHCK).
RN [390]
RN PHOSPHORYLATION (BY MHCK).
RN [391]
RN PHOSPHORYLATION (BY MHCK).
RN [392]
RN PHOSPHORYLATION (BY MHCK).
RN [393]
RN PHOSPHORYLATION (BY MHCK).
RN [394]
RN PHOSPHORYLATION (BY MHCK).
RN [395]
RN PHOSPHORYLATION (BY MHCK).
RN [396]
RN PHOSPHORYLATION (BY MHCK).
RN [397]
RN PHOSPHORYLATION (BY MHCK).
RN [398]
RN PHOSPHORYLATION (BY MHCK).
RN [399]
RN PHOSPHORYLATION (BY MHCK).
RN [400]
RN PHOSPHORYLATION (BY MHCK).
RN [401]
RN PHOSPHORYLATION (BY MHCK).
RN [402]
RN PHOSPHORYLATION (BY MHCK).
RN [403]
RN PHOSPHORYLATION (BY MHCK).
RN [404]
RN PHOSPHORYLATION (BY MHCK).
RN [405]
RN PHOSPHORYLATION (BY MHCK).
RN [406]
RN PHOSPHORYLATION (BY MHCK).
RN [407]
RN PHOSPHORYLATION (BY MHCK).
RN [408]
RN PHOSPHORYLATION (BY MHCK).
RN [409]
RN PHOSPHORYLATION (BY MHCK).
RN [410]
RN PHOSPHORYLATION (BY MHCK).
RN [411]
RN PHOSPHORYLATION (BY MHCK).
RN [412]
RN PHOSPHORYLATION (BY MHCK).
RN [413]
RN PHOSPHORYLATION (BY MHCK).
RN [414]
RN PHOSPHORYLATION (BY MHCK).
RN [415]
RN PHOSPHORYLATION (BY MHCK).
RN [416]
RN PHOSPHORYLATION (BY MHCK).
RN [417]
RN PHOSPHORYLATION (BY MHCK).
RN [418]
RN PHOSPHORYLATION (BY MHCK).
RN [419]
RN PHOSPHORYLATION (BY MHCK).
RN [420]
RN PHOSPHORYLATION (BY MHCK).
RN [421]
RN PHOSPHORYLATION (BY MHCK).
RN [422]
RN PHOSPHORYLATION (BY MHCK).
RN [423]
RN PHOSPHORYLATION (BY MHCK).
RN [424]
RN PHOSPHORYLATION (BY MHCK).
RN [425]
RN PHOSPHORYLATION (BY MHCK).


```

FT STRAND 253 261
FT HELIX 265 268
FT TURN 269 269
FT TURN 273 274
FT STRAND 278 278
FT HELIX 279 287
FT HELIX 290 296
FT HELIX 297 297
FT HELIX 301 303
FT TURN 305 307
FT TURN 316 317
FT HELIX 320 334
FT TURN 335 335
FT HELIX 338 355
FT TURN 356 356
FT STRAND 360 360
FT STRAND 368 368
FT HELIX 373 382

```

Query Match 5.2%; Score 171.5; DB 1; Length 2116;
 Best Local Similarity 20.3%; Pred. No. 0.19; Indels 173; Gaps 25;
 Matches 114; Conservative 101; Mismatches 173;

```

QY 156 DVEDVDVNASEENDSDALDDSDDLAS--D-----YSDVDSQSHSGSRKQNFKKFG 209
DB 983 EVELSEFSESESKDKGVLEKTRVRLQSELDLTVRLDSTKXKSELRLQKKLEEE--- 1039
QY 210 SLDSLSIEQINEPQRMHCPACONGCAIDWNLHPLIARFKGARV---KLHRELAE 266
DB 1040 -----LKQVQ-----ALAETAARKLQEAANKKLQSEYTE 1070
QY 267 VLEK---DLQMRGASVPGGEIYGQWKGLEDKDYETVWPWYITMNTRLDKDDDKML 323
DB 1071 LNEKFSSEVAR-----SNVKSKTLE---SQIVAVNELDESK--- 1108
QY 324 GMSNOELLEFDYKYLRRAR-----HSYGPQGHGMSVLMFESSATGYLAEHLRE 375
DB 1109 ---NRDALR---KKKKLDAWLEMKQLESTG-----GEKKSLYDLKQESDWEALRNQ 1158
QY 376 LAEM-----GLDRIANGQKSMFSGYRQLYFLATQDDIDINQSHQKTRKLKFLKSY 430
DB 1159 ISELQSTIANLEKTI-----KSTLEGEVARLQGELEAEQLAKSNVEKQKKVEYLDLEDKSA 1213
QY 431 Q-----EMVVKELRQ-ISEDNQOLNPFKNLKSQNKAKVLEES-----LEIMS 473
DB 1214 QLAETRAKQALDKLKKLQEQELSEVOTQJSEANNKAVNSDSTNKLHETSFPNKLLELEA 1273
QY 474 EKLRTAEADRIVRORTKQH-----EQNREMDAHDHFPMDSIKQIHERRD----- 520
DB 1274 EQAKAKALEKKRGLGSELKLVNVEQLEEEKKESNEKRVLDLEKVELKQIEEVSAS 1333
QY 521 -----AKENFEMLQOQERAKVYGQOQONIN-----PSSNDCKRAEVSSEFIEF 566
DB 1334 KKAIVTEAKNKESELLEIKRQ-YADVVSRRKSEVEQLKTLQALQNEELRNFAEAEQQLDR 1392
QY 567 QEK-----EMSEFV-----EEREMLIKQO-----EKMEQDK----- 593
DB 1393 AERSKKAEFDLEAVANLEETRAKKAKEMKKAETDYRSTKSELDDAKVNSSEQYVQ 1452
QY 594 -KRHEEIFDLKEFDEALEQ 613
DB 1453 IKRLNELSELRLVLEADER 1473

```

RESULT 12
 MY4 RABIT
 ID MY4 RABIT STANDARD; PRT; 1938 AA.
 AC 028641;
 DT 16-OCT-2001 (Rel. 40, Last Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, skeletal muscle, juvenile.
 OS *Oryctolagus cuniculus* (Rabbit).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
TX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
RA Maeda K., Hostinova E., Roesch-Kleinhauf A., Schuster H., Gasperik J.,
RA Wittnghofer A.;
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit
RT skeletal muscle and a novel cosynthesis of S-1 fragment with the
RT essential and regulatory light chains."
RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Muscle contraction.
CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
CC and 2 regulatory light chain subunits (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
CC meromyosin (LM) and 1 heavy meromyosin (HMM). It can later be
CC split further into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32574; AAA74199.1; -
DR PIR; A59293; A59293.
DR HSSP; P13538; 2MYS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PR02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT DOMAIN 1 783
FT MYOSIN HEAD-LIKE.
FT DOMAIN 842 813
FT COILED COIL (POTENTIAL).
FT NP_BIND 179 186
FT ATP (POTENTIAL).
FT DOMAIN 658 680
FT ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774
FT ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35
FT METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130
FT METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552
FT METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756
FT METHYLATION (TRI-) (BY SIMILARITY).
SQ
SEQUENCE 1938 AA; 223064 MW; DBA8A2EC5B182626 CRC64;

```

Query Match 5.1%; Score 169; DB 1; Length 1938;
 Best Local Similarity 21.1%; Pred. No. 0.23;
 Matches 121; Conservative 94; Mismatches 183; Indels 175; Gaps 28;

```

QY 152 QEPVEDVDVNASEEND-----SDALDS--DDDLASDDYSDVDSQSHSGSRKQNF-- 202
DB 1309 QAFTOQIEELKQLEERIKAKALAHVLAQSAHSDCLLRQEQYERQEAELQRAWMSKAN 1368

```

```

QY 203 -----WFKKFGSISLSISQINEPQRQWHPACQNGPAGIDWNLHPLAARTKGRAR 257
DB 1369 SEVAQMRTRYK-----ETDAIQRTSE-----LEBAKKKLAQR 1399
QY 258 VKIARELAETLENDLQMGASVYPOGEIYGOMKKGDEKDYIWPVWIMNTR----- 313
DB 1400 LQDAEHEVEAVN-----AKCSLEKTKQRLQNEVEDL-----MIDVERINACA 1443
QY 314 -----LJKDNDKWLGMNOBLEEYFPKYEALPARHSYQPGQHGMSVLMF-----ESS 362
DB 1444 ALKKQRN-----FDKILAEWKHKYETTHALEASQKESRSISTEVEFKVKNAYEESLDQL 1498
QY 363 ATGYLEAERLHRELAENGELDRIMWGOR-----SMFGGYRQLYGFL-ARKQDLIDINQHS 417
DB 1499 ETIKREKKNLQOEISDL-TEQIABGGKRIHELEKVKYQVEKSESLQALAEASLEHE 1557
QY 418 QGR-TRAKPEFLKSYQGVENVKELQISEDNOQLNVPKMKLSKONKAKVLESLEISSEKL 476
DB 1558 EGKILRIQLELNQVKESEIDRKIKAEKDEIDOL-----KRI-HIRVESHQSTLDAEI 1608
QY 477 RRTAEDNRIVROTKMQHEQNRREEMDAH--DRPFMSIK-----QIH----- 516
DB 1609 R--SRNDARIRIKKMGEDLNEMEIQLNHNAMAAELRVYNTQGLKDTQLHLDALR 1665
QY 517 -----ERR-----DAK-BENFEMTQQOERAVVQOQ-----ONIN 546
DB 1666 QOEDLKEQLAMVERANMLQAEIHELRLATLEQTERSRAKVAQELDASERVQLHTON-- 1723
QY 547 PSSNDCKRAEVSFIEFOEKEMEEFVE-----REMLIKQOEKK 588
DB 1724 -TSLINKKKKLETTISQI--QGEMEDIQDARAAEKAKKAITDAMMAEELKKEQDTS 1779
QY 589 --MEDMKRHHHEIFDLKEFDEALQOLMYNG 619
DB 1780 AHEEMKKNQYVADLQHRIDEA-EQALTKGG 1811

```

```

CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation. (Probable).
CC -1- SUBUNIT: Monomer. (Probable).
CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC entirely alpha-helical, configured as a series of peptide repeats
CC of varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIF directly by ionic interactions.
CC Domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC different species.
CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC EMBL: L09190; AAA65582.1; -.
CC DR PIR: A45973; A45973.
CC DR HSSP: P02633; 4ICB.
CC DR Genew: HGNC:11791; THH.
CC DR MIM: 190370; -.
CC DR GO: GO:0005856; C:cytoskeleton; NAS.
CC DR GO: GO:0005509; F:calcium ion binding; TAS.
CC DR InterPro: IPR001751; CABP_S100.
CC DR InterPro: IPR002048; EF-hand.
CC DR InterPro: IPR002017; Spectrin.
CC DR Pfam: PF00036; ehfhand_1.
CC DR Pfam: PF01023; S_100; 1.
CC DR ProDom: PD003407; CABP_S100; 1.
CC DR PROSITE: PS00018; EF HAND; 1.
CC DR PROSITE: PS00303; S100 CABP; 1.
CC KM Keratinization; Calcium-binding. Repeat; Citrullination.
CC FT DOMAIN 1 91
CC FT CA BIND 22 33
CC FT CA BIND 62 73
CC FT DOMAIN 314 390
CC FT REPEAT 314 326
CC FT REPEAT 327 339
CC FT REPEAT 340 351
CC FT REPEAT 352 364
CC FT REPEAT 365 377
CC FT REPEAT 378 390
CC FT REPEAT 391 444
CC FT REPEAT 391 444
CC FT REPEAT 391 444
CC FT REPEAT 391 444
CC FT REPEAT 397 402
CC FT REPEAT 397 402
CC FT REPEAT 403 408
CC FT REPEAT 409 414
CC FT REPEAT 415 420
CC FT REPEAT 421 426
CC FT REPEAT 427 432
CC FT REPEAT 432 438
CC FT REPEAT 433 438
CC FT REPEAT 439 444
CC FT REPEAT 444 702
CC FT DOMAIN 923 1162
CC FT REPEAT 923 952
CC FT REPEAT 953 982

```

```

FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1794 1801 QERDROYR -> RESETSTG (IN REF. 2).
FT CONFLICT 1857 1857 O -> K (IN REF. 2).
FT CONFLICT 1880 1880 V -> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MW; A7AB5947FB62B31D CRC64;

Query Match 5.1%; Score 167.5; DB 1; Length 1898;
Best Local Similarity 20.9%; Pred. No. 0.26;
Matches 98; Conservative 85; Mismatches 162; Indels 123; Gaps 22;

QY 246 LLAHARTKARVYKHELAELVEKDIQ-----NRGASVIPCBIYQWKGLEDEKDY 300
DB 1140 LLREERK-RRQELERQYRE--EELQQRKQRYDEDDQSDLKQW-----EPEKNA 1192
QY 301 IWPBWIIMTRLDKNDKMLGMNGOLLEYDKTEALR-----ARSHYGPQHNG-- 353
DB 1193 V-----RNNKYCKGRENEQFRQLEDQVRDQSQODLQHLGEGQERDRE 1238
QY 354 MSYLFESSATGYLEARLHELAWGLDRIAMQO-----KRSMPGSGVRQ 399
DB 1239 QERRRMOQANRHPPEEQLEREBEQKAKRDRKSQBEKQLRBEREKRRQETDRKFRE 1298
QY 400 LYGLATKODLIDFNQ-----HSQKTRLFKFE-----LKSQYEMV 435
DB 1299 EEOQLQREBQPLLRQERDRKFRREELHGOQGRKFLBEOQLREBERKFLKEEQQLRL 1358
QY 436 KEIRQISED-----NOQLN-----YFNKLSKQNKAKVLESLIMSEQLR 478
DB 1359 EEBEQQLAQDRKFRBEEQQLSRQERDRKFRBEEQQVRQERERKFLBEEQQLRQERHRK 1418
QY 479 TAEENRIVRQRTKMO--HEQNRREEMDAHDFPMDSIKOI--HERPDAK-----EEN 525
DB 1419 FRBEEQLQREBQQLHRQER-----DRKFLBEEQQLRQERDRKFRBEOQLNSQBEK 1472
QY 526 F-----EMLQOQERAKVVGQOQONINPSSNDCRKAEEVSSFLFEQKEMEFPVEEREM 581
DB 1473 FLBEEQQLHRQQRKFLBEEQQLRQERQ--QRRQQRDRKFRF--EEOQLRQREBQQLS 1529
QY 582 IKQOEK--MEDMKRHEEFLDEKEFDALBQMLTKG---LHND 624
DB 1530 ROERDRKFRLEQKVRROEQ---ERKMEDEQQLRQERQOQQLROED 1573

RESULT 14
MYHD HUMAN STANDARD; PRT; 1938 AA.
AC GDUKX3; O95252;
DT 16-OCT-2001 (Ref. 40, Created)
DT 16-OCT-2001 (Ref. 40, Last sequence update)
DT 15-MAR-2004 (Ref. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, extraocular (MYH3-ec).
GN MYH3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraocular muscle;
RA MEDLINE=99318869; PubMed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75 (1999).
[2]
RP SEQUENCE OF 1917-1938 FROM N.A.

```

```

RC TISSUE=Extraocular muscle;
RX MEDLINE=99026150; PubMed=9806854;
RA Winters L.M., Briggs M.M., Schachar F.;
RT "The human extraocular muscle myosin heavy chain gene (MYH13) maps to
the cluster of fast and developmental myosin genes on chromosome 17.";
RL Genomics 54:188-189 (1998).
CC -1- FUNCTION: Muscle contraction.
CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
split further into 2 globular subfragments (S1) and 1 rod-shaped
subfragment (S2).
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF111782; AAD29948.1; -.
DR EMBL; AF075248; AAC83241.1; -.
DR HSSP; P13538; 2MYS.
DR Genew; HGNC:7571; MYH13.
DR MIM; 603487; -.
DR GO; GO:0005859; C:muscle myosin; TAS.
DR GO; GO:0003779; F:actin binding; NAS.
DR GO; GO:0005524; F:ATP binding; NAS.
DR GO; GO:0005516; F:calmodulin binding; NAS.
DR GO; GO:0003776; F:muscle motor activity; TAS.
DR GO; GO:0006936; P:muscle contraction; TAS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR Pfam; PF00612; IQ_2.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; myosin_N; 1.
DR Pfam; PF01576; myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR Prodom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ_1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ_1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 COILED COIL (POTENTIAL).
FT NP BIND 843 1938 ATP (POTENTIAL).
FT DOMAIN 179 186 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT MOD RES 761 775 METHYLATION (TR1-) (POTENTIAL).
FT MOD RES 130 130 METHYLATION (TR1-) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381C05 CRC64;

Query Match 5.0%; Score 166.5; DB 1; Length 1938;
Best Local Similarity 19.1%; Pred. No. 0.31;
Matches 118; Conservative 98; Mismatches 204; Indels 199; Gaps 26;

QY 115 SRKYD-----NMFVAPPVSRPPLGGMNQARGSSAGHTVQEPDVEDVDVNASEEB 168
DB 552 NKLYDQHLGKSNFQKPK-----AKGVAERHFSIVHAGTVD-----Y 590
QY 169 NDSDALDSDDDLAS-----DDYDSVDSQKSHGRKQNKWFKKFGS 210

```

```

Db 591 NIAGWIDKKKDDPNETVGLYQKSSILKLSFLPSNYAAGETGDSGGKGGKSSPOT 650
Qy 211 LQSLSTIEQINBP---RQWH-----CPACONGGAIDWY-NLHPLAARTGARRV 259
Db 651 VSAVFPENLAKMTNLRSHPHFRVRLINETKTPGVMDHYLVHQLKNGVLEG---IR 707
Qy 260 LHRE--LAEYLEKOLQNR---GASVIFGCEI-----YQGWK- 290
Db 708 ICKKGFPSRLIYADPFQRRYRIINAASAIPEGQFIDSKNASEKLNSIDVDRQFRGNKV 767
Qy 291 -----GLGEDEKDYIWPVWIIIMNR-----LDKDN-----D 320
Db 768 PFPAAGLLGLEBMRDEKLV---TMTSTQACRGYIMEVEFEKQMERDRSIFCIQYINR 823
Qy 321 KWLGMGNOELLEYPDKYEAALPAHSHYGPQCHRGMSVLM--FESSAGYLEARLELAE 378
Db 824 SFMNVKMPMWNLEFFKIKPLIK---SAEAKEMATMKEDFEKTELARSEKRELEE 879
Qy 379 MGLDRIAMGQKSMFSGGVQLYGFATKQDLDI-----FNHSGQKTRLPFEL 427
Db 880 -----KMSLLQEKNDLQLOQVSETEENLMDAEERCEGLIKSKILL 919
Qy 428 KSYQEMVVELRQISEDNQOLYFENKLSKONKHAKEYLESLEIMSEKLRTEMDNRIVR 487
Db 920 ---EAKYKELTERLEEEEM-----NSELVAKKRLBDCSSLRDIDDELTL- 964
Qy 488 QRTKQHOENRENDNAHFRFMDISIKOIHERRDAKEENFELQOERAKVGGQOONINP 547
Db 965 -LTLYEKKKHAITE-----NKXKNISEMTLLEENISLTJTEKSKLOEAMQO----- 1009
Qy 548 SSNDQCRKRAEVSFSFI-----EFQEKMEEFVEEREMLIKDOE---KXMEEMKRNHE 598
Db 1010 -TLDDLOVEHEDKYNGLIKNALEQOTDLSGLEOEKRLADDERAKRLEGLDKVSOE 1068
Qy 599 EIFDLEKEFEDEALBOUMTK 617
Db 1069 SIMDLENEKQIEBKXKK 1087

RESULT 15
HMMR_HUMAN STANDARD; PRT; 724 AA.
ID HMMR_HUMAN
AC 075330; Q92767;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hyaluronan mediated motility receptor (intracellular hyaluronic acid binding protein) (receptor for hyaluronan-mediated motility) (CD168 antigen).
GN HMMR OR IHABP OR RHAMM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC TISSUE=Breast carcinoma; PubMed=9601098;
RX MEDLINE=96264864; PubMed=9601098;
RA Asmann V., Marshall J.F., Fleber C., Hofmann M., Hart I.R.;
RT "The human hyaluronan receptor RHAMM is expressed as an intracellular protein in breast cancer cells."
RT J. Cell Sci. 111:1685-1694(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=97045829; PubMed=8890751;
RA Wang C., Entwistle J., Hou G., Li Q., Turley E.A.;
RT "The characterization of a human RHAMM cDNA: conservation of the hyaluronan-binding domains."
RT Gene 174:299-306(1996).
CC -!- FUNCTION: Involved in cell motility. When hyaluronan binds to HMMR, the phosphorylation of a number of proteins, including the

```

```

CC focal adhesion kinase occurs. May also be involved in cellular
CC transformation and metastasis formation, and in regulating
CC extracellular-regulated kinase (ERK) activity.
CC -!- SUBUNIT: Subunit of the HARC complex.
CC -!- SUBCELLULAR LOCATION: Cell surface and cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=O75330-1; Sequence=Displayed;
CC Name=B;
CC IsoId=O75330-2; Sequence=VSP 004286;
CC -!- TISSUE SPECIFICITY: Expressed in breast cancer cell lines and in normal breast tissue.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 2.76-84(2001);
CC WWW=http://www.ncbi.nlm.nih.gov/row/guide/80286666_g.htm".
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
CC EMBL: AF032862; AAC32548.1;
CC EMBL: U29343; AAC52049.1;
CC Genew: HGNC:5012; HMMR.
CC MIM: 600936;
DR Hyaluronan acid; Alternative splicing; Repeat; Glycoprotein; Antigen.
KW DOMAIN
FT 635
FT 645
FT 657
FT 666
FT CARBOHYD 133
FT CARBOHYD 133
FT CARBOHYD 477
FT CARBOHYD 477
FT CARBOHYD 567
FT CARBOHYD 567
FT CARBOHYD 588
FT CARBOHYD 588
FT VARSPLIC 90
FT FTID=VSP 004286.
FT CONFLICT 75
FT CONFLICT 103
FT CONFLICT 277
FT CONFLICT 298
FT CONFLICT 322
FT CONFLICT 330
FT CONFLICT 332
SQ SEQUENCE 724 AA; 84031 MW; EA68AD6D2A626926 CEC64;

Query Match 5.0%; Score 165; DB 1; Length 724;
Best Local Similarity 19.0%; Pseq. No. 0.12; Indels 102; Gaps 18;
Matches 91; Conservative 100; Mismatches 186;

Qy 210 LQSLSTIEQINBP---RQWH-----CPACONGGAIDWY-NLHPLAARTGARRVYLHRELAETL- 268
Db 2 SFPAKPKFRNDS-----GAPSPGAYDVLTLEVLKGPVSFKQRFKQKSKONLN 55
Qy 269 -EKDLOW-RGASVIFGCEIYQGMKGLGEDEKDYIWPVWIIIMNRTRDKDNDKMLMG 326
Db 56 VDKQTLTPASARKVKSSES-----KESQKNDKULKLEIKRILVLDQERGAQDSR----- 104
Qy 327 NOELLEIFDKYF-----ALPAHSHYGPQCHRGMSVLMFESSATGYLEART-----HREL 376
Db 105 IQDLETELEKEMARLNAALREKSL-ANNATLEKQILTLTNNELKSKFSENGNOVL 163
Qy 377 AEMGLDRIAMGQKSMFSGGVQLYGFATKQDLIFNQ-----HSGCK----- 420
Db 164 RILSLLEMLKLRNKE-----TKWRGMMAKQEMENKLTQVTRSLSESGKLAQLEGKIV 217
Qy 421 -TRIKELKSYQMVVYKELRQISEDNQOLYFENKLSKONKHAKE-----VLEE 467
Db 218 SIEKELKDEKSEKLELEYEISCSADQVEKYLDAQLEHKLKKNDEILSKQSLSE 277
Qy 468 SLEIMSEK-----LRTADNRIVRQRTKQHOENREMDAHDREFMDSTIQIHE 517
Db 278 NIVILSKQVEVDLNVKQULEKEKEDH--VVRNREHNNELNAEMQNLKQKFLB--QOERE 333

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: April 6, 2004, 19:30:55 ; Search time 45 Seconds
(without alignments)
4382.197 Million cell updates/sec

Title: US-10-030-829-3
Perfect score: 3313
Sequence: 1 MSSRGPMSKXKXVQGGYRP.....EFDEALEQLMKYKGLHNEED 625

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3313	100.0	625	10	Q9LDX1	Q9LDX1 arabidopsis
2	778	23.5	304	10	Q7XY17	Q7XY17 triticum ae
3	548.5	16.6	170	10	Q7XYE6	Q7XYE6 triticum ae
4	242	7.3	647	10	Q8VZ79	Q8VZ79 arabidopsis
5	233.5	7.0	644	10	Q9SMN2	Q9SMN2 arabidopsis
6	215.5	6.5	629	10	Q9SBW2	Q9SBW2 oryza sativ
7	211	6.4	634	10	Q9SA11	Q9SA11 arabidopsis
8	208	6.3	1003	12	Q91LK9	Q91LK9 kaposi's sa
9	203	6.1	628	10	Q81S39	Q81S39 zea mays (m
10	198	6.0	1036	12	Q9DUM3	Q9DUM3 kaposi's sa
11	194.5	5.9	447	10	Q8H8B2	Q8H8B2 oryza sativ
12	192	5.8	634	10	Q9S9P3	Q9S9P3 arabidopsis
13	192	5.8	2760	5	Q815Y2	Q815Y2 plasmodium
14	190	5.7	635	10	Q9C7B0	Q9C7B0 arabidopsis
15	190	5.7	639	10	Q8W563	Q8W563 triticum mo
16	189.5	5.7	638	10	Q91HB1	Q91HB1 arabidopsis

17	187	5.6	694	10	Q9AX36	Q9AX36 oryza sativ
18	186.5	5.6	554	10	Q9M150	Q9M150 arabidopsis
19	186	5.6	761	13	P79793	P79793 gallus gall
20	186	5.6	1941	13	Q8UWA0	Q8UWA0 gallus gall
21	183.5	5.5	561	10	Q91TU0	Q91TU0 arabidopsis
22	183	5.5	559	10	Q04614	Q04614 arabidopsis
23	181	5.5	976	12	Q9DUN0	Q9DUN0 kaposi's sa
24	180.5	5.4	895	10	Q48878	Q48878 sorghum bic
25	180	5.4	662	10	Q23064	Q23064 arabidopsis
26	179.5	5.4	1388	6	Q28021	Q28021 bos taurus
27	179	5.4	1777	5	Q813P4	Q813P4 plasmodium
28	178.5	5.4	1957	5	Q04010	Q04010 onchocerca
29	178	5.4	985	5	Q9U0S5	Q9U0S5 mytilus gal
30	178	5.4	1705	5	Q9U0S7	Q9U0S7 mytilus gal
31	177	5.3	915	5	Q869R0	Q869R0 dictyostel
32	177	5.3	1129	12	Q9GR71	Q9GR71 kaposi's sa
33	176	5.3	1089	12	Q40947	Q40947 kaposi's sa
34	176	5.3	1464	5	Q811F6	Q811F6 plasmodium
35	175.5	5.3	1503	5	Q861Q0	Q861Q0 dictyostel
36	174	5.3	1375	13	Q90Y37	Q90Y37 brachydanio
37	174	5.3	1939	6	Q9TV63	Q9TV63 sus scrofa
38	173	5.2	488	5	Q45420	Q45420 caenorhabdi
39	172.5	5.2	607	11	Q8BS10	Q8BS10 mus musculu
40	172.5	5.2	1156	16	Q66878	Q66878 aquilex aeo
41	172.5	5.2	1871	10	Q9SRD5	Q9SRD5 arabidopsis
42	172	5.2	892	11	Q63939	Q63939 rattus sp.
43	171.5	5.2	1391	11	Q922J3	Q922J3 mus musculu
44	171.5	5.2	1939	6	Q9TV61	Q9TV61 sus scrofa
45	171.5	5.2	1946	5	Q97291	Q97291 plasmodium

ALIGNMENTS

RESULT 1	ID	Q9LDX1	PRELIMINARY;	PRT;	625 AA.
AC	Q9LDX1	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DE	01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
DT	EMBL CAB62356.1 (SGS3) (Hypothetical protein).				
GN	AT5G23570.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=COLUMBIA;				
RX	MEDLINE=20181125; PubMed=10718197;				
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,				
RA	Tabata S.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence				
RT	features of the regions of 3,076,755 bp covered by sixty P1 and TAC				
RT	clones.";				
RL	DNA Res. 7:31-63(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=20306668; PubMed=10850495;				
RA	Mourrain P., Beclin C., Bimayan T., Feuerbach F., Godon C.,				
RA	Morel J.-B., Jonette D., Lacombe A.-M., Nikic S., Picault N.,				
RA	Remoue K., Sanial M., Vo T.-A., Vaucheret H.;				
RT	"Arabidopsis SGS2 and SGS3 genes are Required for Posttranscriptional				
RT	Gene Silencing and Natural Virus Resistance.";				
RL	Cell 101:533-542(2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Yamada K., Chan M.-M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,				
RA	Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,				
RA	Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,				

RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Arabidopsis Full Length cDNA Clones",
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones",
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB025633; BAA97244.1; -
 DR EMBL; AF339719; AAF73360.1; -
 DR EMBL; BT002944; AA022757.1; -
 DR EMBL; BT004380; AA042374.1; -
 DR InterPro; IPR005380; XS.
 DR InterPro; IPR005381; zf.
 DR Pfam; PF03468; XS; 1.
 DR Pfam; PF03470; zf-XS; 1.
 DR Hypothetical protein;
 SQ SEQUENCE 625 AA; 71971 MW; 456E2A1396706A96 CRC64;
 Query Match 100.0%; Score 3313; DB 10; Length 625;
 Best Local Similarity 100.0%; Pred. No. 1,7e-189; Indels 0; Gaps 0;
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSSRAGPMKSKKNVGGYRPEVEQVVOGLAGTSLASSDGDGEWEVISKKNKPKNTSG 60
 DB 1 MSSRAGPMKSKKNVGGYRPEVEQVVOGLAGTSLASSDGDGEWEVISKKNKPKNTSG 60
 QY 61 KTWVSQNSNPRAWGQGGQGGGNSVSGGNNVSGGNGNGGICQANISGGRALSRKYDN 120
 DB 61 KTWVSQNSNPRAWGQGGQGGGNSVSGGNNVSGGNGNGGICQANISGGRALSRKYDN 120
 QY 121 NFVAPPPVPRPLLEGMMWQANGSAGHTAVOEFPDVEDVDNASEENDSDALDDSD 180
 DB 121 NFVAPPPVPRPLLEGMMWQANGSAGHTAVOEFPDVEDVDNASEENDSDALDDSD 180
 QY 121 NFVAPPPVPRPLLEGMMWQANGSAGHTAVOEFPDVEDVDNASEENDSDALDDSD 180
 DB 121 NFVAPPPVPRPLLEGMMWQANGSAGHTAVOEFPDVEDVDNASEENDSDALDDSD 180
 QY 181 LASDDVDSDVSKSHSRKQNKPKFKGSLDLSISIQINERQOMHCPACQNGGALDW 240
 DB 181 LASDDVDSDVSKSHSRKQNKPKFKGSLDLSISIQINERQOMHCPACQNGGALDW 240
 QY 181 LASDDVDSDVSKSHSRKQNKPKFKGSLDLSISIQINERQOMHCPACQNGGALDW 240
 DB 181 LASDDVDSDVSKSHSRKQNKPKFKGSLDLSISIQINERQOMHCPACQNGGALDW 240
 QY 241 YNLHPLLAHARTGARRVYLHRELAVLEKDLQMGASVTPCGEITYGQMGIGDEKDY 300
 DB 241 YNLHPLLAHARTGARRVYLHRELAVLEKDLQMGASVTPCGEITYGQMGIGDEKDY 300
 QY 241 YNLHPLLAHARTGARRVYLHRELAVLEKDLQMGASVTPCGEITYGQMGIGDEKDY 300
 DB 241 YNLHPLLAHARTGARRVYLHRELAVLEKDLQMGASVTPCGEITYGQMGIGDEKDY 300
 QY 301 IYWPBPVYIINWTRLDKDDNDKVLGMGNOELLEFDYKTEALRRHSYSGRHSVLMFE 360
 DB 301 IYWPBPVYIINWTRLDKDDNDKVLGMGNOELLEFDYKTEALRRHSYSGRHSVLMFE 360
 QY 301 IYWPBPVYIINWTRLDKDDNDKVLGMGNOELLEFDYKTEALRRHSYSGRHSVLMFE 360
 DB 301 IYWPBPVYIINWTRLDKDDNDKVLGMGNOELLEFDYKTEALRRHSYSGRHSVLMFE 360
 QY 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGVYOLYGLAKODLDLFNHSSQK 420
 DB 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGVYOLYGLAKODLDLFNHSSQK 420
 QY 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGVYOLYGLAKODLDLFNHSSQK 420
 DB 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGVYOLYGLAKODLDLFNHSSQK 420
 QY 421 TRLFELKSYOEWVVKELROISENOQJNFKNLSQNHAVLESLIEMSEKLRTA 480
 DB 421 TRLFELKSYOEWVVKELROISENOQJNFKNLSQNHAVLESLIEMSEKLRTA 480
 QY 421 TRLFELKSYOEWVVKELROISENOQJNFKNLSQNHAVLESLIEMSEKLRTA 480
 DB 421 TRLFELKSYOEWVVKELROISENOQJNFKNLSQNHAVLESLIEMSEKLRTA 480
 QY 481 EDNRIVQRTQMOHEONREMDADRFPMDSIKOIHERRPAKEENPEMLQOORAVVGO 540
 DB 481 EDNRIVQRTQMOHEONREMDADRFPMDSIKOIHERRPAKEENPEMLQOORAVVGO 540
 QY 481 EDNRIVQRTQMOHEONREMDADRFPMDSIKOIHERRPAKEENPEMLQOORAVVGO 540
 DB 481 EDNRIVQRTQMOHEONREMDADRFPMDSIKOIHERRPAKEENPEMLQOORAVVGO 540
 QY 541 QOONINPSNDCKRAEVSSEFEPOKEMEVEVEREMLIKDOEKMEDMKRHHET 600
 DB 541 QOONINPSNDCKRAEVSSEFEPOKEMEVEVEREMLIKDOEKMEDMKRHHET 600
 QY 541 QOONINPSNDCKRAEVSSEFEPOKEMEVEVEREMLIKDOEKMEDMKRHHET 600
 DB 541 QOONINPSNDCKRAEVSSEFEPOKEMEVEVEREMLIKDOEKMEDMKRHHET 600
 QY 601 FDLEKPEDEALQOLMYKHGLNEDD 625
 DB 601 FDLEKPEDEALQOLMYKHGLNEDD 625

DB 601 FDLEKPEDEALQOLMYKHGLNEDD 625
 RESULT 2
 Q7XY17 PRELIMINARY; PRT; 304 AA.
 ID Q7XY17
 AC Q7XY17
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE EMBL.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OC NCBI_Taxid=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. PH 82-2-2;
 RA Zhao X., Li Q., Zhang X.;
 RT "Isolation and expression of a new kind of gene involve in
 RT embryogenesis in Triticum aestivum L.",
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF542974; AAP80862.1; -
 DR EMBL; AF542974; AAP80862.1; -
 SQ SEQUENCE 304 AA; 35958 MW; 89980215A9584208 CRC64;
 Query Match 23.5%; Score 778; DB 10; Length 304;
 Best Local Similarity 51.2%; Pred. No. 9,7e-39;
 Matches 151; Conservative 82; Indels 4; Gaps 2;
 QY 325 MGNQELLEFPDKYALRRHSYSGRHSVLMFESSATGYLAERLHRELAEWGLDRI 384
 DB 1 MGNQELLEFPDYATATARRHAYGPGGRGMSVLIFFSSAYGYAEERLHGHFDORTDRD 60
 QY 385 AMGQKSMF-SGVYOLYGLATKODLDIFNHSQGRTRKFLKSYOEWVVKELROI 443
 DB 61 TWQNRVPRPLLEGMMWQANGSAGHTAVOEFPDVEDVDNASEENDSDALDDSD 120
 QY 444 DNQOLNFKNLSKONKAKVLESLEIEMSEKLRTADNRIVQRTQMOHEONREMDA 503
 DB 121 DNQOLNFKNLSKONKAKVLESLEIEMSEKLRTADNRIVQRTQMOHEONREMDA 180
 QY 504 HDRFPMDSIKOIHERRPAKEENPEMLQOORAVVGOQOONINPSNDCKRAEVSSE 563
 DB 181 OERFHDQLENIHKTEDSEFEFLQEBRAKA---RQCDVDSGTENRRLRKEVORF 237
 QY 564 IEFQKEMEVEVEREMLIKDOEKMEDMKRHHETFDLEKPEDEALQOLMYKH 618
 DB 238 IECVQDVQEFERENDEMIKAHEKKVQKKEYMAKEVELEKPEDEALQOLMYKH 292
 RESULT 3
 Q7XY17 PRELIMINARY; PRT; 170 AA.
 ID Q7XY17
 AC Q7XY17
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE HORR (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OC NCBI_Taxid=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. PH 82-2-2;
 RA Li J.R., Wang F., Li Q.Z., Zhang X.S.;
 RT "Gene isolation and expression of a new Zn-finger",
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF469493; AAP80610.1; -
 SQ NON_TER 1

SQ SEQUENCE 170 AA; 19980 MW; 0AFD21F178B3068 CRC64;
Query Match 16.6%; Score 548.5; DB 10; Length 170;
Best Local Similarity 59.8%; Pred. No. 2.2e-25;
Matches 101; Conservative 31; Mismatches 36; Indels 1; Gaps 1;
QY 322 WLMGNOELLEYPDKTEALRARRSHYSGPOGRGMSVLMFESSATGYLAEERLHRELAEMGL 381
DB 1 WLMGNOELLEYPDSVAATARARAYGPGHGRGMSVLMFESSAGVMEAEELHNFIDORT 60
QY 382 DRANQKSMF--SGGYRQLYGFLATKODLDINQHSQGRTRLKFKELKSYQEVVYKLRQ 440
DB DRPTWQNRVRVPLPGGRQLYGFLARKEDMETNRHCQGRSLKYEWRSHNEVVAQMKQ 120
QY 441 ISEDNQOLNFKKLSKQNHAKVLEESLEIMSEKLRRTAEDNRIVRQR 489
DB 121 MSEDNQOLNFKKLVKVTQORSKVETLGVITQKLRTEWENIYVRSK 169
RESULT 4
Q8VZ79 PRELIMINARY; PRT; 647 AA.
AC Q8VZ79;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN AT3G48670.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh U., Carninci P., Chen H.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh U., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamlaya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY065184; AAL38360.1; -;
RA EMBL: BT000136; AAN15455.1; -;
DR EMBL: IPR005379; XH.
DR InterPro: IPR005380; XS.
DR InterPro: IPR005381; zF.
DR Pfam: PF03469; XH; 1.
DR Pfam: PF03468; XS; 1.
DR Pfam: PF03470; zF-XS; 1.
KW Hypothetical protein.
SQ SEQUENCE 647 AA; 74824 MW; 4AE83BE6376279D0 CRC64;
Query Match 7.3%; Score 242; DB 10; Length 647;
Best Local Similarity 23.1%; Pred. No. 2.3e-06;
Matches 111; Conservative 89; Mismatches 193; Indels 88; Gaps 19;
QY 181 LASDDYDVSQKSHSRKQKMFKKFPGSLDLSIEQINEPQRMHCPACONGPAIDW 240
DB 7 LSSDDSDSDISSE-----EMDEYGDKMVYLNKGGKLVKRLSPQ-AFICPCPNKKKTSFQ 60
QY 241 YNLHPLAARFTYG-----ARRVKLRELAELVEKDIQMGKASVIPCGETIYGQMKL 292
DB 61 YK--DLLQHASGVGNSNDKRSKAKERASHLALVKYLQDILADASAEAPSSKQKGNPF 118

QY 293 GEDEKDEIYVPMVILNTRLDKDDNDKWLGMGNOLLE-EPDK-YEALRARRSHYSGPOG 350
DB 119 QDDDHBEKLVYPMKGLVNNIPPTKADGRSAGSSSKLDEYILRQFNPRVPLNNYIG 178
QY 351 HRGMSVLMFESSATGYLAEERLHRELAEMGLDIRIANGQKRSMSGGVRLYGLATKODL 410
DB 179 HSGTAIVEFKDNGNLGNGLLPDKAYTVDHGKKDKLK---DGRKLGLYGTARADY 234
QY 411 ---DINQHSQGRTRKLF-----ELKSYQEVVYKLRQISEDNQOLNFKKLSKQNH 461
DB 235 NGNNIIGENIRKTDGKTALTEEARQOELLVQNLRLVER-----KKKD 281
QY 462 AKVLEESLEIMSEKLRRTAEDNRIVRQRTQMGRSEMDADRPFMDSIKOIHERRDA 521
DB 282 MKETFEELCVSKSEL-----NQLMEKGRNQKQHR-ELNAIQETMTSHIOCTIVD---- 330
QY 522 KEENFEMLOQOERAKV-----GOOQONINSSNDCKRAEVSSTLEPOE 568
DB 331 DHEKLRLESERKLEIKCNELAKREVNNGTERML-----SEDLQNSKNS-LELA 385
QY 569 KEMEERFEREMLIKQOEKMEKRRHHEIFDLEKDE-----ALEQL-----MYK 617
DB 386 MEQKADDEEVKCLAEQORQKEEL---HEKILRLRORDQKQALFEVEQLKQGLNVK 441
QY 618 H 618
DB 442 H 442
RESULT 5
Q9SMN2 PRELIMINARY; PRT; 644 AA.
AC Q9SMN2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN TAB19.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choinsene N., Robert C., Brottier P., Wincker P., Catolico L.,
RA Artiguenave F., Saurin F., Weissenbach J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Queflier F., Salanoubat M.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL133315; CAB62356.1; -;
DR PIR: T46211; T46211.
DR InterPro: IPR005379; XH.
DR InterPro: IPR005380; XS.
DR InterPro: IPR005381; zF.
DR Pfam: PF03469; XH; 1.
DR Pfam: PF03468; XS; 1.
DR Pfam: PF03470; zF-XS; 1.
KW Hypothetical protein.
SQ SEQUENCE 644 AA; 74411 MW; AFS5AAD6F9811EB7 CRC64;
Query Match 7.0%; Score 233.5; DB 10; Length 644;
Best Local Similarity 22.9%; Pred. No. 7.4e-06;
Matches 110; Conservative 88; Mismatches 192; Indels 91; Gaps 19;
QY 181 LASDDYDVSQKSHSRKQKMFKKFPGSLDLSIEQINEPQRMHCPACONGPAIDW 240
DB 7 LSSDDSDSDISSE-----EMDEYGDKMVYLNKGGKLVKRLSPQ-AFICPCPNKKKTSFQ 60

QY 241 YNLPLLAHARTKG-----ARRVKLHRELAEVLEKDLQMRGASVIPCSEIYGQWKL 292
 Db 61 YK--DLLQHSAGVGNNSNDKRSKAKESHIALVYLIQDLADSASAEFSSKQKNGFI 118
 QY 293 GEDDEKYEIWPMPVIIMNTRLDKDNDKMLGNGOELL-EYFDK-YEALRAHSHYSGOG 350
 Db 119 QGDHDEKLVYPMKGIIVNIPPTKADGRSAGSGSKLDEYILRGFNPTRVPLNNYLG 178
 QY 351 HNGMSVIMPESSATYGLAEERLHRELAEVLEKDLQMRGASVIPCSEIYGQWKL 410
 Db 179 HSGTAIVERKQWNGHNLGLFDKAYTVDGHGKMDLKK---DGRKLGXYGIARADY 234
 QY 411 ---DINQHSOGKRLKF-----ELKSQOEMVYKELROISEDNQOLNFKLISKONH 461
 Db 235 NGNNIIGENDIRKTGDKTLTAELTEESARKOELLVQRLQVLE-----KKKD 281
 QY 462 AKYLESELEIMSEKLRRTAEDNRIVRQRTQMOHQREEMADHREFMDSIKOIHERRDA 521
 Db 282 MKEIEELGVSKSEL-----NQLMEKEKNQOKHYR-ELNAIQERTMGIQKIVD---- 330
 QY 522 KEENFMLOOQERAKYV-----GQOQONINSSNDCKRAEVSFTIEFQ 568
 Db 331 DHEKRLLESEKRLKIKCNELAKREVNHTERMKL---SEDLQONSKNS-LELA 385
 QY 569 KEMEEFVEEREMLIKQEKKMDMKRHHHEIFDLKEPDE-----ALEOL---MYK 617
 Db 386 MEQOKADEEVKCLADEOREL-----HEKIRLERQDQKQALIEVLEQLKGLNVK 438
 QY 618 H 618
 Db 439 H 439

RESULT 6

Q9SBW2 PRELIMINARY; PRT; 629 AA.

AC Q9SBW2; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, last annotation update)
 DE Putative transcription factor X1.
 GN P0663E10.7.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriharitideae; Oryzaceae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Lemont;
 RA Chen M., Lucas J.R., Bennett J.L.;
 RT "Different rates of divergence in SH2/Al-homologous regions of rice."
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF101045; AAF21887.1; -
 DR EMBL: AP004317; BAB90725.1; -
 DR Gramene; Q9SBW2; -
 DR InterPro; IPR005379; XH.
 DR InterPro; IPR005380; XS.
 DR InterPro; IPR005381; zf.
 DR Pfam; PR03469; XH; 1.
 DR Pfam; PR03468; XS; 1.
 DR Pfam; PR03470; zf-XS; 1.
 SQ SEQUENCE 629 AA; 72715 MW; 4BB7928A7E3BF4F1 CRC64;

Query Match 6.54; Score 215.5; DB 10; Length 629;
 Best Local Similarity 21.1%; Pred. No. 8.5e-05;

Matches 110; Conservative 96; Mismatches 179; Indels 137; Gaps 23;
 QY 183 SDDVSDVYQSHSGRKQKMKFKFEGSLDLSIEQINEPQOMHPACQNPALIMWN 242
 Db 4 SDEDESEISDS-----EIDEDVADKFPYARL-VAGFPKYVDGSGYSCPL-SKKKKD-FN 54
 QY 243 LHPLLAHTKGA---RRVK--LHRELAEVLEKDL-----QMRGASVIPCSEIYGQWKL 290
 Db 55 INNLIQHSAGVGAABNRQAKDKATRALAKHKNLGLTKSSGQSGQTAVER----- 105
 QY 291 GEDDEKYEIWPMPVIIMNTRLDKDNDKMLGNGOELL-EYFDK-YEALRAHSHYSGOG 350
 Db 106 -OPLPRDEKFEVPMWGVLVNPTWKDG-RQIGSGNHLKEQLSRFCPLKIIPLMNFPG 163
 QY 351 HNGMSVIMPESSATYGLAEERLHRELAEVLEKDLQMRGASVIPCSEIYGQWKL 409
 Db 164 HSGNAIVERKQWNGHNLGLFDKAYTVDGHGKMDLKK---DGRKLGXYGIARADY 219
 QY 410 -----LDIFNQHSG----- 418
 Db 220 TSPGLIGHLRKNGDKLTINDENEGARKTDLVANLANQIEVKNRHQLELVYNERTT 279
 QY 419 -----GK-----TLKELKSQOEMVYKELROISEDNQOL-NYRKNLISKONKAKYLE 466
 Db 280 SLEKMGQREQLQKYNNEIRKMQQLAQHSQKIIDENQKLSLESKSELSINTRSKEID 339
 QY 467 ESELEIMSEKLR-----TAEDNRIVRQRTQMOHQREEMADHREF 507
 Db 340 E-IAKSDYDRRIIDQOKNAIKSSHLKTLATLEQERADENVLKVREKREKKA- 394
 QY 508 FMDSIKQIHERDAAKEENFMLOOERAKYVGOQOONINSSNDCKR-RAEVSFTIEF 566
 Db 395 -YKLIKLEQGVDAKQ-LELDIQLKGLYVMKIM---PQEDSALKKIKIDELSE-BL 447
 QY 567 QEK--EMEEFVEEREMLIKQEKKMDMKRHHHEIFDL 603
 Db 448 QEKMDLDMESLNQTLVYKERKSNTEMQDARKELENGLLD 489

RESULT 7

Q9SA11 PRELIMINARY; PRT; 634 AA.

AC Q9SA11; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, last annotation update)
 DE F23A5.14 protein.
 DE F23A5.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S., Lee J.M., Li J.,
 RA Gonzalez A., Liu A., Liu K., Sakano H., Koo T., Pham P., Vayenberg M.,
 RA Howing B., Chin C., Choi E., Chlou J., Alatafi H., Brooks S., Chao Q.,
 RA Conn L., Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C.,
 RA Lam B., Nguyen M., Palm C., Shinn P., Tambunga G., Davis R.W.,
 RA Ecker J.R., Federgruel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F23A5 sequence."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF004317; BAB90725.1; -
 DR Gramene; Q9SA11; -
 DR InterPro; IPR005379; XH.
 DR InterPro; IPR005380; XS.

Dp	Db	Oy	Oy	Oy	Oy	Oy	Oy	Oy
152	QEFPEVEDVDNASEENDSDALDDSDDLASDPYDSDGKSKSGSKKKMKFAGGSGL	211	:	:	:	:	:	:
394	EEDDEEDEDDEEDEDEBEDDEDDEDDEDEDEDEDEDEDEDEDEDEGGDKNK-----TL	446	:	:	:	:	:	:
212	DSLSIQINEPQRQ--WHCPACONGPFGAIDWYNLHPILAHARTGARRVYLHRE-LAEV	267	:	:	:	:	:	:
447	SIGSSQQQEPPQOEFPQOGEFPQOOEPPOQGEFFLPQPQGEFFPQOGEFPQOGEFP	506	:	:	:	:	:	:
268	LEKDLOMRGASVIPCEGIYGWKLGEGEKDYEIWPMPWIINATRLDKDNNDKLGMGN	327	:	:	:	:	:	:
507	LQRPQOEPRQGEFRPQOGEFPQOGEFPQODQOOD-----EQODEQQODEQQODE	556	:	:	:	:	:	:
328	QELLEYFDKYEAIRAHNSYGRPGHRMSVLTPRESSATGYLEARHLHRELAEWGDRIAWG	387	:	:	:	:	:	:
557	QQODEPQODEQQODEPQODEQQODEPQODEPQODEPQODEPQODEPQODEPQODE	608	:	:	:	:	:	:
388	QKSMFGSVRQLYGFPLATKODIDIFNGHSOGKTRIKLFELKSYYQMVVYKLRQISBDNQ	447	:	:	:	:	:	:
609	-----EQELEBEOQ-----ELLEBOELEBOELEBOELEBOE	639	:	:	:	:	:	:
448	IYTFANKLSKONKHAKYLESLSEIMSEKLRRTADENRIVROSTKNQHECNREMDANDRF	507	:	:	:	:	:	:
640	LEBOQELEBOQELEBOQELEBOQEDEL-----EBEQELEBOQELEBOQELEBOQE	695	:	:	:	:	:	:
508	FMDTSIKQHNERDAKENPFEMLOQERAKVVGQQQNINPNSSNDCRKRAEVSSFIEQ	567	:	:	:	:	:	:
696	LEEQELEBEBOELEBOQELEBOQELEBOQELEBOQELEVEBOQEBOQEBOELEB-VEEQ	754	:	:	:	:	:	:
568	EKEMEEFYE-BREMLIKQOEKMCKRHNEEIFDLKEFDLAESOL-----MYCG	618	:	:	:	:	:	:
755	EEQEBOQEBOELEVEBOQEBOELEVEBOQBEOELEVEBOEQGVGOQEOTVEBFIIIH	814	:	:	:	:	:	:
Dp	Db	Oy	Oy	Oy	Oy	Oy	Oy	Oy
619	GJHNEDD	625	:	:	:	:	:	:
815	GSSSEDE	821	:	:	:	:	:	:
RESULT 11								
Q8HBZ	PRELIMINARY;	PRT:	447 AA.					
ID Q8HBZ								
AC Q8HBZ	Created)							
DT 01-MAR-2003	(TReMBIrel).	23,	Last sequence update)					
DT 01-MAR-2003	(TReMBIrel).	23,	Last annotation update)					
DT 01-JUN-2003	(TReMBIrel).	24,	Last annotation update)					
DM Hypothetical protein.								
GN OUI13AF05.17.								
OS Oryza sativa (japonica cultivar-group).								
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
CC Spermatophyta; Magnoliopsida; Liliopsidae; Poales; Poaceae;								
OC Eubacteroidae; Oryzaceae; Oryza.								
OX NCBI_TaxId=39947;								
RN [1]								
RP SEQUENCE FROM N.A.								
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,								
CU Currie J., Coltura K.; "Rice Genomic Sequence."								
RT Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.								
LT DR EMBL; AC099401; AANO6846.1; -								
DR InterPro; IPRO05380; XS.								
DR Pfam; PF03468; XS; 1.								
KW Hypochetrical protein.								
KW SEQUENCE 447 AA; 49300 MW; CADE079CDE4857B7 CRC64;								
Query Match								
Best local similarity		5.9%; Score 194.5;	DB 10; Length 447;					
Matches 62; Conservative 36; Mismatches 76;								
Indels 65; Gaps 7								
Oy	228 CPAONGCALDWNYLNPHLAHAAT-----KGARVYLHRELAEVLEKDLQMGASVIPFC	282	:	:	:	:	:	:
Db	223 CLAC--GRSSXDFADFAHGVLNHAINPNWDSPFDIGHGLHKALCVLMGW----	269	:	:	:	:	:	:
Oy	283 GEIYGOMKGLGEDKDY-----IWPMPWIINATRLDKDNNDKMLGMGN	328	:	:	:	:	:	:

Db 270 -----YTKVPENSKAYQSLPDLVQASREDLIIMPTVIIHNTATGKRRKKGRAEGLGNK 323
QY 329 ELLEFYDK-----YEALRARSYSYGPQHRGMSVLMESSATGYLAEARLHRELAEWGLD 382
Db 324 EM---DKTISELGFAGSKSKSLYKRGHGLTLIFANSPLAGLKEERLADTLERODHG 379
QY 383 RIAMGQRSMFSGGV-----RQLYGFLATKODLDIFNCHSGKTRLK 424
Db 380 RIGWLMARANQSGVSNPLVETDNRGTGKERLLIVGYLALISSDMELDSRKRASLX 438

RESULT 12

Q9S9P3 PRELIMINARY; PRT; 634 AA.
AC Q9S9P3;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE T24D18.1 protein.
GN T24D18.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lenz C., Pham P.,
RA Toriumi M., Chin C., Chlou J., Choi E., Chung M., Gonzalez A.,
RA Hwang B., Koo T., Li J., Liu A., Vayberg M., Altafi H., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,
RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010924; AAF18488.1; -.
DR PIR; E86293; E86293.
DR InterPro; IPR005379; XH.
DR InterPro; IPR005380; XS.
DR InterPro; IPR005381; ZF.
DR Pfam; PF03469; XH; 1.
DR Pfam; PF03468; XS; 1.
DR Pfam; PF03470; ZF-XS; 1.
SQ SEQUENCE 634 AA; 72635 MW; 0071930DED7F41B0 CRC64;

Query Match 5.8%; Score 192; DB 10; Length 634;

Best Local Similarity 22.1%; Pred. No. 0.0022;

Matches 98; Conservative 80; Mismatches 179; Indels 86; Gaps 20;

QY 225 ONHCPCQNGRPAIDWYNLHPLAHA-----RTGARRVYKJHRELAEVLEKDLQMRGA 277
Db 38 QLRCPCCA-GKKKQD-YKTKELYAHTGVSKGSATSLQKANHLLALAMELEL-AGYA 94
QY 278 SVLPCEIYGQWKGLEDEKDYEIWPPVVIINNTRLDKDNDKMLGMGNQELLEIFYDKX 337
Db 95 ELPVRFPPVVRPQDTEPBNPHNYVWPWVGIVVPL-KEADKELLDSAYVLTLSK 152
QY 338 EALRAHSHSGPQHRGMSVLMFESSATGYLAEARLHRELAEWGLDRIAMGQRSMFSG-G 336
Db 153 KPLEVNAFWEDOSIVGLAKFGDWSGFAGATELKEFEFTQSSKKEWTER----SGDS 208
QY 397 VROLYGFATKODL-----DIFNCHSGKTRLKFEELKSYOE--WVYKELRQI-----SE 443
Db 209 ESKAYGWCARADPFESQSPIGETLSKEGQLRTSDISQKNVQDRNVTLELSIMIAMTNE 268
QY 444 DNGQNLVFNKLS-----KQNGKAVLESLSEIMSEKLRRTAEDNRIVQRTKMOR 494

Db 269 DLNKQVSYNRTAMSLQRLVDEKKNLHOAFDETKKQOMSLNHI---OKILYKEKELSN 325
QY 495 EONREEMDAHDFPMDSIQIHERRDAPKEENFEMLOOQRAKAVVGQOONINPSSNDPCR 554
Db 326 ELDRKRRDESR-----AKOL-----EKHEALTELDROKL-----DEBK 359
QY 555 KRAEVSFFIEFOEKEMEBFEVEREMLIKOEKMEKRRHHEELFDEKEFD--EAL 612
Db 360 RKSDDANKSLQLASRRQKADSSVLTVEBHQKQEDALNK-----ILLLEKQDLYQTLE 415
QY 613 -----QLMYKRGHLEHEDD 625
Db 416 MEIOELKRGKLOVM-KH-LGDDDD 436

RESULT 13

Q815Y2 PRELIMINARY; PRT; 2760 AA.
ID Q815Y2
AC Q815Y2;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL0315C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.-J., Sun B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carnucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";
RL Nature 419:498-511 (2002).
DR EMBL; AE014845; AAN36152.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2760 AA; 330896 MW; EE9964C4845181AC CRC64;

Query Match 5.8%; Score 192; DB 5; Length 2760;

Best Local Similarity 19.7%; Pred. No. 0.014;

Matches 136; Conservative 142; Mismatches 272; Indels 142; Gaps 32;

QY 10 KEKNVGGYRPEYEQVQGLAGTRLASSODDGEWEVISKKN-KKKPGNTSGKTWVSQN 67
Db 1473 KENDVNSKNERIEIVQVNEIKVTN--NKEEBG-----KKKILKEKEINDCLNDYINQ 1524
QY 68 SNPPRAMGGQOQGRGNSVGRGNVSGRGIOANISGRALSRKYONNFPAPPE 127
Db 1525 K-----KKEKKNNMAMTGRPIYKRONRNINIKNDLK--KYSSKSESGF----- 1568
QY 128 VSRPLEGGMNQARGSQAHTAVQEPFVEDVDN-ASEEENDSDALDSD-----DDL 181
Db 1569 -----NDYAFALAEFFVITGVNSEPDYLSIDNQANNEKHNDIINNHNITIKSKK 1621
QY 182 ASDDYDSD-----VSQKSGSKKQNKWPFKFGSLDSLSEIQ 219
Db 1622 KENIYNSPHTYGRPIYKSKSNPNYNNKIKSTTHNALIKKRRKKTINKSINSFTK 1681
QY 220 NEPQRO-WHCPCQNGRPAIDWYNLHPL-LAA-----ARTGARRVYKJHREL----- 264
Db 1682 NSNANKIVRTSKA--NTIDYNNSTIKITIHKEQNVDDQGYIDLTKRKLIYDALDEI 1739
QY 265 -AEVLEKDLQMGASVIPCGEIYQ-----WKGLEDEKDYEIWPPVVIINNTRLDKD 318

DB 1740 NCOTQXNKLKIT-ENITGVGKKGQNVENIKNCTG-----AMLEIKTRKGN 1787
 QY 319 NDKWLMGNOEL-----LEVPKYEALRBHSGYQGRHMSVLMFESSATGYLEAE 370
 DB 1788 NEB--DQFSEBELKALEKLLKELKRIEELKLEBEKKRLLEENSLAVEBEKKKKK 1845
 QY 371 RLHR--ELAMGIDRLTAMGQKSMFSGVRLYGF--LATKODDIPNOHQ--GTRLEK 425
 DB 1846 EMKMEEMKIELQKKEEKLQYSRQEQIRKLEELKKEKLELLEEMNLQEQKRE 1905
 QY 426 ELKSYQEMVYKE--LRQISEDNQOLAFKXKLSKONKHAFLYEESELSKELRTIADNR 484
 DB 1906 EIKMEBEKREKELKIKKEQK--KEEMKQEQKQK--KQKIKM--EEMKKEEQK 1957
 QY 485 IYRQRTMGHEONREEMDAHREFMDSIKQIHERDAKEMENMLQOQRAVYQOQON 544
 DB 1958 KVELK--LEBEKKREELK-----LEBQKREERKMEBEKKREBEKKEELIKMEBE 2011
 QY 545 INPSSNDCKRAEVSFIEFOEKME--FVBERMLIKQDEKMEKMK--KRHE 598
 DB 2012 IKKMEBE--KKKKEBEKKMEBEKKKEBEKKKEBEKKKEBEKKKEBEKKKEBE 2069
 QY 599 EIPDLE-----KEFDEALEQMLYQGLHN 622
 DB 2070 EIKFLEBEKKEHKKIYENKENTSEKOLLHN 2101

RESULT 14
 09C7B0 PRELIMINARY; PRT; 635 AA.
 AC 09C7B0; PRELIMINARY; PRT; 635 AA.
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)
 DE Hypochemical protein.
 GN T2E22.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unseid M.,
 RA Fairman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delaney M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choiane N., Artiguenave F., Robert C., Brotier P.,
 RA Wincker P., Catolico L., Weisenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Huer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Dronek H., Erfle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Reichelt J., Scharte M., Schoen O., Bargues M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Orlowski B., Clément D.,
 RA Cooke R., Lande M., Berger-Liauro C., Puntieri B., Maury D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Collet A., Casacuberta E.,
 RA Monfort A., Argitton A., Flores M., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mannheim G., Haase D., Schoof H., Liguori R., Vitale D.,
 RA Mayer K.F.X., Kahl S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
 RA Cressy T.H., Haas B., Malt R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Milišević J., Sellers P., Gill J.B., Feldblyum T.V.,
 RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asanuma E.,
 RA Saemundsson S., Kohara M., Matsuno M., Matsuno A., Muraki A.,
 RA Kiyokawa S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana.";

RL Nature 408:820-822 (2000).
 DR EMBL: AC069474; AAG51004.1;
 DR InterPro: IPR005379; XS.
 DR InterPro: IPR005380; XS.
 DR InterPro: IPR005381; zF.
 DR Pfam: PF03469; XH; 1.
 DR Pfam: PF03468; XH; 1.
 DR Pfam: PF03470; zF-XS; 1.
 DR Hypothetical protein.
 KM SEQUENCE 635 AA; 73841 MW; 9C738BFD87DBE8A3 CRC64;
 SQ

Query Match 5.7%; Score 190; DB 10; Length 635;
 Best Local Similarity 21.4%; Pred. No. 0.0029;
 Matches 96; Conservative 87; Mismatches 198; Indels 68; Gaps 18;

QY 197 SRQKRWFKKFFSGSLDSLEIQINERQRMHCPAQNGAIDWNLPLAHARTKAR 256
 DB 2 NKLSDPEKLLYKLLSGKLE-VKVSYRFLCPDKNKKVGLY--VDILQASGVNS 58
 QY 257 RYK-----LHRELAEVLEKDLQMRGASVI-----PCGE---TYGMKGL 292
 DB 59 QSKRSILTEKASHALAKLILKDLAHYATSTISKRLKARTSFIPATGDAPIY----- 112
 QY 293 GEDEKVEIYWPVWVIMN--TRLDKDNNDKMLGNGOELLEFDK--YEALRBHSGYQ 349
 DB 113 -DQAQPEKLVWPMKGLVNIPTSTEDGRCSTGESGPKLDELIRGFNPVRVYTWDRF 171
 QY 350 GHRGKSVLMFESSATGYLEAERLHRELAEWGLDRIAMGQKSMFSGVRLYGFATKOD 409
 DB 172 GHSQGLYEFNRDNGQDLVFKKAVEGQ-----HGKDMVCGATDSLSVWMLNAD- 225
 QY 410 LDIPNOHQKTRLEK-ELKSYQEMVYKELRQISEDNQOLNTR--XKLSKQK--HKVLE 466
 DB 226 -DYVANTLGNLKMGPDKSYRFAEEARKKQKQLQRIINFEVENQYKLLQIKYSQ 284
 QY 467 ESLIEMSEKLRTADNRIVRQRTMGHEONREEMDAHREFMDSIKQIHERDAKEMEN 526
 DB 285 DEVKIKYE---TEKEKILNAYSBDLTGRQKQTDIFNNIFND---HEKQVQLS- 334
 QY 527 EMLQOQRAVYQOQONINPSSNDCKRAE--VSSFIEFOEKMEBEFVEREMLIK 593
 DB 335 -QIKELFIRKLELARBENETQRTVAKLEQNAINISVQLSALQOQKTRQKRLAV 393
 QY 584 DOEKKMEKRRHHEIFDLEKEDEALE 612
 DB 394 DHEKEL-----HKRIALERQDQKQ 415

RESULT 15
 08W563 PRELIMINARY; PRT; 639 AA.
 AC 08W563;
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, last annotation update)
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
 DE Putative transcription factor XI.
 OS Triticum monococcum (Binkum wheat) (Small spelt).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticaceae; Triticum.
 OC NCBI_TaxID=4568;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21898330; PubMed=11901130;
 RX Li W., Gill B.S.;
 RT "The collinearity of the Sh2/Al orthologous region in rice, sorghum and
 RT maize is interrupted and accompanied by genome expansion in the
 RT Triticaceae.";
 RT Genetics 160:1153-1162 (2002).
 RL EMBL: AF434705; AAL5831.2;
 DR InterPro: IPR005379; XH.
 DR InterPro: IPR005380; XS.
 DR InterPro: IPR005381; zF.

This Page Blank (uspto)


```

Db      680 EFLDNDADATLKLRLQAQELARQAQAIKILRLQKQEQARVAKKQAQIIMAAEE----- 734
Qy      414 NQHSQCKTLKFKELK-SYQEMVVKELRQISEDN-----QQNLVFNKXLSKQNKHAKVLEES 468
Db      735 -----KQKQEKQIKIMKQEKIKRIQQRIMEKELRAQOILKAKKKKEEAMANKLLEAE 788
Qy      469 LEIMSEKLRRTAEDNRIVQRITKMOHQREEMDADRPFMDISIKQIHERDAKEENFEM 528
Db      789 KRICEKEMR-----QOAVLKH-QERERRRQ-----MLMLAMERKKAEEK--ER 833
Qy      529 LOQERAKVVGQOQONINPSSNDCKRAEBVVSFLFEQKEMEFEVEREMLIKQDEKK 588
Db      834 LKQEK-----DEKLN-----KERTLEQRLELEW-AKELKKP 866
Qy      589 MEDM 592
Db      867 NEDM 870

RESULT 2
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwanatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1

```

Query Match

5.4%; Score 179.5; DB 2; Length 1388;

```

Best Local Similarity 20.8%; Pred. No. 1,86-05;
Matches 116; Conservative 108; Mismatches 182; Indels 151; Gaps 27;

Qy      147 QHTAVPEPPDVEDVDVNASEENDSDALDDSDLDASDDVSDVSQSHSGRKQNKMKFK 206
Db      509 QHMAVEYQKADHEADKRNLENDVNSIKDQLEDIK-----KRNQ----- 549
Qy      207 FFGSLDLSLEQINERQRMHCPACONGPAIDVNLPLHAHTKGR----- 256
Db      550 -----SQISTEKVNLQRO-----LDEN--ALRTESDTRARLKTQAESSK 590
Qy      257 -----RVKLHRE--LAEVLEKIQMR--GASVPCGEIYGOM 289
Db      591 QIQQLSNNRDLDQKNCLETAKLKEEFIMLOSVLSESRDRTHGSEII--NDLQRI 648
Qy      290 KGLGEDEKDYIYVPMVLIIN-----TRLDKNDQKILGWG-----NOELLEYPDKY 337
Db      649 SGLBEDVYKNGKILLAVYLEKRLQERFTDLKEKKNNEIDMTYQTKYIQSLBGEETEH 708
Qy      338 EALRABHSYGPQGRGWSVLMFESSATGYLAEERL--HRELAEWGLDRIAMGQKR-SMF 393
Db      709 KATYARLADQKTIYESIE-----EAKSEAMKEMEKLESERTIKQKVENILLBAERCSIL 764
Qy      394 SGGRVQLYGLATYKQDI-----DIFNQHSQCKTLKFKELSYQEMVVKELRQISEDNQ 446
Db      765 DCDLKQ-----SQQKINELLKQKQDVLENDVRLT-LKTBQETQKCLIQ--NDLKMQTQ 815
Qy      447 QLAYFK--NKLKQNGHAKVLEESLEIMSEKLRRTAEDNRIVQRITKMOHQREEMDA 503
Db      816 QVNTLKSEKQKQENHLEMKSLKQNAELRKE-----ROADQCMKELQPOLBA 868
Qy      504 HDRF--FMDSIKQIHERDAKEENFEMLOQ-----EAKVVGQOQONINSSND-C 553
Db      869 EGYFSTLYKTQVRRLKECEKTKLCKELQKQELQDERDSLAQLEITLTADSEQLA 928
Qy      554 KRAAEVSSFLFEQKEMEFEVEREMLIKQDEK-----KMEDMKKRHHHEIFDL- 603
Db      929 RSIABEYSDLE-KEKTKKE-LEIKEMARKQBLTEKDATIASLEETNTLTSDVANLA 986
Qy      604 -EKE-----FDEALEQL 614
Db      987 NEKEELNKKLKEAQEQL 1003

RESULT 3
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-418-710-72

Query Match 5.2%; Score 173.5; DB 4; Length 1969;
Best Local Similarity 20.1%; Pred. No. 8,6e-05;
Matches 128; Conservative 94; Mismatches 213; Indels 203; Gaps 25;
Qy      47 ISKKNKPKGNTSGKTWVSQNSNPRAWGQOQGRGSNTVSGRGNVSGRGNNG----- 100

```

```

Db 341 LSTGRRFFGN---QTPWPMASPLI-----HQGKKAVSNVNVPKYQHHHSPLIVE 393
Qy 101 --RGIQANISGRGALSRKYNDNFVAPPVSRPPEEGNMWQARGSAAHTVAOEPDVE 158
Db 394 QFRGTDSDI-----PSSK-----DSEDSNEDEEDDEE 421
Qy 159 DDVUNASSEENDSDALDSS-----DDDLASDDYDSVDQKSHGSKKQKMKF 205
Db 422 EDEEDEDSDSDSQSEDSNSSESDTEGSEEDDDDKDQDESDSD---TEGEXTSMKLNK 477
Qy 206 KFFGSLDLSISIQINBPORQWCHPACQONPGAIIDWNLPLI-----AHARTKG 254
Db 478 TTSSKSPMSLGHSTPRYLHIAKAPGAPALCSCSGPAPFGTSSSTLTSPPHGTSK 537
Qy 255 ARVYLHRELEVLKED-----LQMGASVIPCGEYIGW-----KGL 292
Db 538 RRVYDERBELRPLFEGWQRETRINFGRLQGEVAVYAPCGKMLQYEVIKYLSRNGI 597
Qy 293 GEDEKD-----YELVW-----PBYVITWNLTKDDNDKWLG 324
Db 598 MDISGDNFSFSAKIRVGDPEYARDGPOEQWCLLKEEDVIPIRAMEGRRGPNNP--- 653
Qy 325 MGNQELLEYFDKY---BALRABSYGPOGHGWSVLMFSSA---TQYLEABRLHRELAEM 379
Db 654 -----DQARABESRMRRRKRRPPNVGNAEFLDNADALRLKLOQEA-RQAAQI 702
Qy 380 GLDIRIAMQKTSFSGVQOLYGLATKODLDI FNOHSGQKTRLKFKELSY-QEAWVKEL 438
Db 703 KLRRLQOQEOQARVAKKQAQAIMAAE-----KKQKQKQIKIHQKQEKIKRI 751
Qy 439 QOISEDN---OQNLVFNKLSKQNKHAYLEBSLEITSEKLRRTAEDNRIVYQRTKMOH 494
Db 752 QOIRKEKELRAQOILLEAKKKKEEBANAKLLEAKRIKEREMR-----RQAAVLTK 802
Qy 495 EONREMAHDFPFMDSIQIHERDAKENEPMQQOERAKVYGOOQONINSSNDRC 554
Db 803 ROERRRRROH---MMLMAMEARKKAEBK--ERLKQER-----DEKRLN----- 842
Qy 555 KRAEVSFFIEQKEMEFEVEREMILIDQEKMEEM 592
Db 843 -----KERKLEBRLELEN-ALELKKPNEDM 867

RESULT 4
US-08-056-200-94
: Sequence 94, Application US/08056200
: Patent No. 5616500
: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-chul
: TITLE OF INVENTION: Trichonyalin and Transglutaminase-3 and
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olason & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/056,200
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Pedrick, Michael F.
: REGISTRATION NUMBER: 36,799
: REFERENCE/DOCKET NUMBER: NIH054.001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (714) 760-0404
: TELEFAX: (714) 760-9502
: INFORMATION FOR SEQ ID NO: 94:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-056-200-94

Query Match      5.1%; Score 167.5; DB 1; Length 1898;
Best Local Similarity 20.9%; Pred. No. 0.00024;
Matches 98; Conservative 85; Mismatches 162; Indels 123; Gaps 22;

QY 246 LLAHATFKGARVYLKHLHLEAVLEKDIQ-----MGASVIRPGETIYQWKGKGEDEKDYE 300
Db 1140 LLREERER-RRQELERQYRE--EEELQQRKQRYDEDDRSLLKQW-----EPKXENA 1192
QY 301 IWPWPVILINNTRIDKQDNDKWLGMNGQELLEYDXYEALR-----ARHSYGPQHG-- 353
Db 1193 V-----RUNKVYCKRENEQFRQLEDQVQDRQSQDILQHLIGQOEKRE 1238
QY 354 MSVIMFESSATGYLAEERLHLEAEMGLDRIAMQ-----KSMFSGGVQ 399
Db 1239 QERRRWQGANRHPFEEBQLEBEQKAKRRDRKSQBEKQLLREERERRRQETDRKRE 1298
QY 400 LYGLFATKQDLDIFNQ-----HSQKTRLKFE-----LKSQYQENV 435
Db 1299 EEOQLQREOPPLRQERDRKFRFEEELHBOQGRKFLFEEQRRLREERERKFLKEEQQLRL 1358
QY 436 KEIQLQIED-----NQQLN-----YFNKLSKQNHAKVLEESLEIMSEKLR 478
Db 1359 EERQQLQDRDRKFRFEEBQLSRQERDRKFRFEEBQYVRQERERKFLFEEBQQLQERHRK 1418
QY 479 TAEINRIYVRORTKQ--HEQNREEMDAHDFMDSIKQI-HERPDAK-----EEN 525
Db 1419 FREEBQLQREBQQLRQER-----DKFLFEEBQLRQERDRKFRFEEBQLRQERPERK 1472
QY 526 F---EMLIQOERAKVVGQOQONINPSSNDCKRAEVSFLFEOKEMEEFVEEREMT 581
Db 1473 FLFEEBQLLHQQRQKFLQEBQQLRQERQ--QRQORDRKFRE--EQQLRQEREBQQLS 1529
QY 582 IKQDEK--MEDMKGRHHEEFLDEKFFDEALBQIMTKG---LHND 624
Db 1530 RQERDRKFLFEEBQKVRQEQ---ERKFMEDQQLRQEBQOQQLROED 1573

RESULT 5
US-08-800-644-94
: Sequence 94, Application US/08800644
: Patent No. 5958752
: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: TITLE OF INVENTION: Park, Sang-Chul
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:

```

Query Match	5.1%	Score 167.5;	DB 2;	Length 1898;
Best Local Similarity	20.9%	Pred. No. 0.00024		
Matches 98;	Conservative 85;	Mismatches 162;	Indels 123;	Gaps 22;

DB 1193 V-----RDNNKYCKGKGENEGCRQLEBSQVRRDSQDDLCHLLGGEQERDRE 1238

[illegible]

DB 1299 EBOLIOEREBOPILNRDRKFRREELHNOEQGRKFLBEEORLREERERKFLKEEQORL 1355

QY 436 KELOAISED-----NQOLN-----YFKNLKSKONKHAYLBESELEIMSEKLR 478

Db 1359 EREDELQRDRKRFREEEQQLSRQEDRKFFREEEQVVRROBERERKFLFEEOQLROERHRK 1416
QY 479 TAEDNRIVRQTKMQ-HEONNEEMDADRFRMDSIKOI-HERRDAR-----EEN 525

```

DB 1419 FREEBOLLOEKEEQUINRGER ---DRK LBBEEWAKRANBNNT NKGJUNOGE 581
QY 526 F---EMLOOQERAKYVGOOONINPSSNDCKRAAEVSSFIEFOEKMEEFVEEREML 581

```

QY 582 IKDOEKK--MEMMKRRHHEIFDLKEKFEDEALGOIWMYKHG---LHNED 624
::: ::
::: :: :

RESULT 6
HC-08 728-333A-2

```

; Sequence 2, Application US/2022/0000000
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan

```

APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.

Query Match	5.0%;	Score 166;	DP 2;	Length 1102;
Best Local Similarity	19.6%;	Pred. NO. 0.00016;		
Matches 46;	Conservative 66;	Mismatches 85;	Indels 38;	Gaps 5

70 : | : : : : | : : : : |
Db 735 QQQQDEQQQDEQQQDEQQEQEEQEQLLEEDDQEQELEEQE 794

```

      : | : |
795 QELBEQEQET-----BEOQELEBEOQELEBEOQELEBEOQELEBEOQE 836
      : | : | : | : | : | : | : | : | : | : | : | : | : |
527 EMT OOOEPAYVGOOONNTNPSSNDDCRKRAEFV---SGPIEFOEKEMEBFVEREMI 582

```

QY	DB
583 KD----	837 QELEQE----
QEKMEDEMKRRHHEIFDLEKEFEALBQ-----	VEEQEQVEVEEQEQEQEQLEVEVEEQEQEQEQEQEQLEVEVEEQEQEQLE 892
MYKGLHNEED 625	

Db 893 EEVEEQEELLEVEEQELLEVEEQELLEVEEQEVEEQEETVEEPILHGSSSE 947

```

US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:

```

APPLICANT: Ballestas, Mary E.
APPLICANT: Kaye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO

```

; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; PUBLISHED APPLICATION NUMBER: US 60/109,422

```

```

; EARLIER FILING DATE: 1998-11-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0

```


QY 575 VEREWLIDKOEKMEKMK-----HHHEIFDLKEFDEPALEQML 615
 DB 1015 DKKQDLKANTQTLTADVDNNEILKQKRIKFILOEKNE-LQOML 1058

RESULT 10
 US-08-685-871-2
 Sequence 2, Application US/08685871
 Patent No. 6013499

GENERAL INFORMATION:

APPLICANT: NARUMIYA, Shuh
 APPLICANT: IWAMATSU, Akihito
 TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,871

FILING DATE: 24-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-184102

FILING DATE: 25-JUN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-262553

FILING DATE: 14-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 16887/845

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1354 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-685-871-2

Query Match

Beet Local Similarity 5.0%; Score 164; DB 3; Length 1354;

Matches 118; Conservative 89; Mismatches 201; Indels 146; Gaps 27;

QY 147 QHTAVQEPVEDVDVNASBEENDSDALDDSDLDASDDVDVDSQKSHG----- 197

DB 492 QHRINRYORKAEQNEKRRNVENVSTLKQLEDLKYSQNSQLAKBKLSQQLKQLEBAN 551

QY 198 -----RKQNTKFKKPGSDLSISTEQINEPQWCHCPACQNGPAGIDPYNH 244

DB 552 DLRTRESTAVRLKSHTEMSKISQLESLN-RELQERNRILLENKSKQTDK--DYVQLQ 607

QY 245 PVL-AHARTK-----ARRVKHLRELAVLELDLQMRGASVAPCEIYGQWKG-- 291

DB 608 ALLEARRDRGHDSMIGDLQARTSLQEEVGH-LKHNLE-----KVGSRKEAQ 656

QY 292 --LGDEDEYELVWPVYIMNTRLDKQNDKMLGMSQELLEYFDYKALRAHSTYGPQ 349

DB 657 DMLNHSKE-----KNLEFDLNTKLSL-QQRLQEGVNEHKVTKALTDKHQ 703

QY 350 G-HRGMSVLMPESSATGYLLEARLHRELAENGDLRIANGQR-SMFGSGVRLQGLFATLK 407

DB 704 SIEAKSVAMCEMK-LKEERAREKA-----NRVQIEKQCSMLDVLQKQ-----SQ 752

QY 408 QDLDFNHSQG-KTRLKFEKSYOEYVYKELRQISEDNQANFKKLSKONKHAYLE 466

DB 753 QKL-----HILTKNKRMEDEVKN-----LTQLQEQ--ESNKL-LLONELTKQAFELNPK 801

QY 467 ESLEINSEKLRRTADNRIVR-----QRTKM-----QHEQNRSEMDADR--FMDSIK 513

DB 802 GLEKQKQKQINLTLEKRLLEFELALTKQYRGNGCQWRELQDQLAEQYFSTLYKTQVK 861

QY 514 QIERRDAK-EENFEMLQ-----OERA 535

DB 862 ELKEBELKNNRELLKQIQLQNEKETLATQDLAETKASSQLARGLEBOYFELTQSK 921

QY 536 KVVQOQONINPESNDCKRAEYVSFIPEQKEMEFEVEREWLIDKOEKMEKMK 595

DB 922 KAASRNQEL--TDKHTVSRLEANSML--TQIETLRNNEILTQMKKAEYVYLE 976

QY 596 HHHEIFDLKEKFE 609

DB 977 KEELISMLKAFK 990

RESULT 11

US-08-685-576-4
 Sequence 4, Application US/08685576
 Patent No. 5906819

GENERAL INFORMATION:

APPLICANT: Kaibuchi, Kozi

APPLICANT: Iwamatsu, Akihito

APPLICANT: Nakano, Takeshi

APPLICANT: Ito, Masaaki

APPLICANT: Takahashi, No. 5906819aki

TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,576

FILING DATE: 24-JUL-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-325129

FILING DATE: 20-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-17150

FILING DATE: 05-JAN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-131206

FILING DATE: 26-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 16887/843

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1388 amino acids

TYPE: amino acid

;; SEQ ID NO 154
;; LENGTH: 966
;; TYPE: PRT
;; ORGANISM: Murine sp.
US-09-291-417D-154

Query Match 4.9%; Score 162; DB 4; Length 966;
Best Local Similarity 18.3%; Pred. No. 0.0026; Indels 256; Gaps 32;
Matches 138; Conservative 121; Mismatches 238;

```

19 REVEEVLQVGLAGTRLASS-----QDDGEMEWI-----47
Db RPSAQLQLHPVSRVTSNKAELVAEAKAEVMEIEDGRDGEDAVIDAAPPVNH 341
Qy 48 -SKKNKPKPNTSGKTWVSNPNPRAWGQOQSGSNVSGKANNVSGNGNGIOAN 106
Db 342 QDSANVTQPLDSNK--LLQDSSSTPLPSQPE-----PVNGPSQPSGDG---PLQTT 390
Qy 107 ISGRGRLSKYDNFAPPPV--SRPLEGGMWMOARGGSAQHTAVOEFPVEDVDNA 164
Db 391 SPADGLS---KNNDLKVPPVPLKSRP-----LSMDAR---IQMDEKQIPD-----431
Qy 165 SEENSDSDALDDSDDLASDDYDSVQSKHSGSKKQKMFKFFGSL---DSLSIQINE 221
Db 432 -QDENPSPA-----ASKSQKANSRPNSSALETLGGEALINGLELSSVT 476
Qy 222 POROMHCACON-----GPAIDWYNLHPLAHARTKARVYLHRELAVLEKOL 272
Db 477 PSHSRKASDCSNLSTSESMYGTSLADSLNKETGSLKGS---KLH-----NKTLL 526
Qy 273 QMRGASVIPCBEI-YGQWKGLEDEKDEYEVMPVITINTRLDKDNQKWLQMGQELL 331
Db 527 KTRRFVVDGVEVSTTSKIISEDEK-----KDEEMFPL-----560
Qy 332 EYFDYELARHSHYGFQGHGCMSTLMESSATGYLAEHLH-----RELAEML 381
Db 561 ---RROELRELRLQKEHNTQTL---SKHELOEOMHKKFEOEINAKKKFYDEL 612
Qy 382 DRIAWQKSM-----FSGGVOLYGLATKODLIFNOSHQKTRLKFKLSYQEMV 435
Db 613 ENIERQKQOQVEKMEBDHVRKREKAKRILEDODRYAKFOEOLK-QMKKVKSEVEKLP 671
Qy 436 KELRQ-----ISPDNQALYFNKULSKQKHAHYLESLTISE 474
Db 672 RQOKESMKQKMEHSHQKORLDRFAKQKEDLELAKLITTEKR-R-FIDCKEDCISK 730
Qy 475 K--LR-----RTAEDNRIVRORTYQO-----HEONREMDAHRFF 508
Db 731 KQELLRREALWEMEHQLOERHQVLVQOLKQVFLQRHDLIRKHEKEREQOMQRYQRM 790
Qy 509 MDSI-----KOIH-----ERRDAKE 524
Db 791 MEQIKVAVQOQEKALPKIQRSDETRWAMYKSLIHNGASASEQREKTKOSQOEKQ 850
Qy 525 NFEMLQOQERAK-----VVGQOQON--INPSNDCKRAEVSFIPOKEEMEEVE 576
Db 851 KAEELQOQKHQEHOMRQVAVQCESNMSELQOLNKKCYLLVEHETQKALADESHQSIK 910
Qy 577 EREMLIKQOEKQMED--MKRRHEEIFDLEKE 606
Db 911 EMRDKLRKKALEEDINQKREQEMFPLSEB 943

```

RESULT 14
US-09-914-259-11
Sequence 11, Application US/09914259
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999

CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 3878
TYPE: PRT
ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 4.9%; Score 161.5; DB 4; Length 3878;
Best Local Similarity 19.0%; Pred. No. 0.002; Indels 179; Gaps 26;
Matches 111; Conservative 101; Mismatches 194;

```

163 NASEENDSDALDDSDDLASDDYDSV-SQSHSGSKKQKMFKFF-----207
Db 1668 NENQVEVEEOTFKKELDRKPEVPELISNERYALQANNRLLITLLEVYKTAVEET 1727
Qy 208 ---FSLDLSIEQINEPOROMHCACONGPAIDWYNLHPLAHARTKARVYK--L 260
Db 1728 IGRVILGILDRSSKQ-----SSASLIW-----RSEKASVYKCV 1762
Qy 261 HRELAVLEKOLQMRGASVIPCBEIYGQWKGLED--EKDYEVMPVITINTRLDKOD 318
Db 1763 HEHTRVTDISITPSYSGSDMPRNDI-NMMSKYTEGETLSQRLVNSGFA--GTEIDPEN 1818
Qy 319 NDKWLMGMN-----QELAEYFDYELARHSHYGFQGHGCMSTLMESS 367
Db 1819 EELMNLISRLQAVKLELLEAISRTS-QLEHAKVTO-----TELRESFROKQKATBSL 1872
Qy 368 EA---ERLHRELAVLEKOLQMRGASVIPCBEIYGQWKGLED--ATQODLIFNOSHQKT 421
Db 1873 KOEELRERLHEE-----SRABQALVLSGAEVIDGYADEKTLFROIQEKT 1921
Qy 422 RLKPELKSQYQEMVYELROISEDNQALYFNKULSKQKHAHYLESLTISE 472
Db 1922 DIIDLEQELLCASTRQELAEBOQIQEERELLSQKEMKQKAEVGEQOLLQETKLM 1981
Qy 473 SEKLRTAEDNR-----VRORTYQOHEONREMDAHRFF 505
Db 1982 KELEVOQCAEKVVRDLOKQVAKLEIDVEBOVSFIPEQKRTMLMDLQONALEKOL 2041
Qy 506 ---RPFMD--SIKQIHERDAKENFEMILQOQERAKVQO-----QOONINSSND-- 551
Db 2042 EKRRKFLDQALDREH-RDVPQOEIQLEEQ--LKVPRPQIPSEHQTRVEQOLANHLK 2098
Qy 552 ---DCKRAEVSFIPOKEEMEEF-----VEER-----578
Db 2099 EKTQKSEILLSKEQLORDIOERNEIERK-LEFRVLEQALLYSADTFQKVEBRKHFGA 2157
Qy 579 ---EMLIKQOEKQMEDMKKHHHEIFDLEKEFDLEQOLMYKH 618
Db 2158 VEAPELSEVLQOERDALDRKEKEITNLEBQLOFPBELEKNK 2202

```

RESULT 15
US-08-938-105-3
Sequence 3, Application US/08938105
GENERAL INFORMATION:
APPLICANT: Weinand, Leslie A.
APPLICANT: Vikstrom, Karen L.
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,105
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Mannell M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3595-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1886 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-105-3

Query Match 4.8%; Score 159.5; DB 4; Length 1886;
Best Local Similarity 19.3%; Pred. No. 0.001;
Matches 112; Conservative 106; Mismatches 193; Indels 169; Gaps 29;

QY 139 WQARGGSAOHTAVOEPFVDVVDVNASEEND--SDALDGS--DDDLASDDYDSVQKSH 195
DB 1248 WQLTGKGLST--QOMEDLGRQLEBGRKAKNALAHQASRHDCLIREQYEEMENKAE 1305
QY 196 GSRKQNK-----WFKKFFGSLDLS--IEQINEPOROWHCPACONGPAIDWYNLHPL 247
DB 1306 LQRLVLSKANSEVAQWRTKY--ETDAIQRTTELEBAKKL--AQRLQDAEAVEAVN----- 1357
QY 248 AHARTGARRVKLHRELAEV--LEKDLQKRGASVIFCGEITYGQWKGIGEDKDYEIWPP 305
DB 1358 --AKCSLEKTK--HRLQNETIEDLMVVERSNAAA-----AALDKKQNFDKI--- 1401
QY 306 MVIMNTRLDKDNNDKMLGNGNOLLEYPDYKALRAHSYGPQGRGMSVLMF----- 359
DB 1402 -----LAWKQKYESQSELESSQKEASLSTELFKLNAY 1437
QY 360 ----ESSATGYLEAERLHRELAEMGLDRIAWGQK-----RSMFGGVQQLYGFATK 407
DB 1438 EESLEHLETFKRNKRLQBERISDL--TQOLGEGGNVHELEKIRKQLEVEKLEQSAL--- 1493
QY 408 QDLDIFNQSQK--TRLKTELKSYQEMVVELRQISENQOLNYFKNKLSKONKAKV-- 464
DB 1494 BEAASLEHHEGKILRAQLEFNQIKAEIERKLAEKDEMEQ-----AKRN--HLRVVD 1544
QY 465 -LEESLEIMSEKIRTAEDNRIYRQRTKQHEQNREEM-----DAH 504
DB 1545 SLQTSID-----AETRSREALRVKKKMGDLNEMETQLSQANRIASEAQKHLKNAQAH 1598
QY 505 DREFFM-----DSIKQ-----IHERRDA-----KEENFEMLOOQERAKVVGQOQ---- 542
DB 1599 LKOTQLOLDAYANDIKENIATIVERNTLLQALEBELRAVVGQTRSRKLAQELIET 1658
QY 543 ---QINIPSSNDDCKRAEEVVSFIEQEKEMEEFYEE-----REML 581
DB 1659 SERVOQLHSQNTSLINQKKMDADLSQLQTEVEEAVQECRNAREKAKAITDAAMMAEEL 1718
QY 582 IKDQEKK--MEDKKRHHETIFDLKEKFPDALQOLMYKHG 619
DB 1719 KKEQDTSAHLEKRNKNEQTIKDLQHRLEDA--EQIALKGG 1757

Search completed: April 6, 2004, 19:36:00
Job time : 26 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:34:56 ; Search time 46 Seconds
(without alignments)
3568.388 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 3313
Sequence: 1 MESSRAGPMSEKKNVQGGYRP.....EPDEALQWYKRLHNEDD 625

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubppa/US06_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubppa/US08_NEW_PUB.pep:*
7: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
12: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
13: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubppa/US10_PUBCOMB.pep:*
17: /cgn2_6/prodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1428.5	43.1	611	US-10-425-114-58390	Sequence 58390, A
2	745.5	22.5	324	US-10-424-599-234018	Sequence 234018, A
3	721	21.8	223	US-10-424-599-234019	Sequence 234019, A
4	297	9.0	188	US-10-424-599-242612	Sequence 242612, A
5	258	7.8	219	US-10-424-599-178379	Sequence 178379, A
6	199.5	6.0	227	US-10-425-114-68998	Sequence 68998, A
7	199.5	6.0	414	US-10-425-114-63998	Sequence 63998, A
8	182	5.5	1972	US-10-425-114-54044	Sequence 54044, A
9	182	5.5	1972	US-09-839-479-21	Sequence 21, Appl
10	178	5.3	1879	US-10-296-115-1265	Sequence 43, Appl
11	175	5.3	1879	US-10-296-115-1265	Sequence 1265, Appl
12	173.5	5.2	1969	US-09-839-479-71	Sequence 71, Appl
13	172.5	5.2	593	US-10-363-616-295	Sequence 295, Appl
14	172.5	5.2	1156	US-10-369-493-43	Sequence 43, Appl
15	172	5.2	892	US-10-205-219-102	Sequence 102, Appl

16	170.5	5.1	676	14	US-10-256-250-16	Sequence 16, Appl
17	170.5	5.1	725	12	US-10-257-377-1	Sequence 1, Appl
18	170.5	5.1	725	14	US-10-256-250-14	Sequence 14, Appl
19	170.5	5.1	725	15	US-10-023-634-78	Sequence 78, Appl
20	170.5	5.1	725	15	US-10-023-634-79	Sequence 79, Appl
21	170.5	5.1	725	15	US-10-205-647A-4	Sequence 4, Appl
22	168	5.1	1043	15	US-10-310-154-449	Sequence 449, Appl
23	167.5	5.1	3899	14	US-10-171-311-4	Sequence 8, Appl
24	167.5	5.1	3917	14	US-10-171-311-8	Sequence 34255, A
25	166	5.0	345	14	US-10-029-386-34255	Sequence 2, Appl
26	166	5.0	1162	12	US-09-894-273-2	Sequence 2, Appl
27	166	5.0	1162	14	US-10-294-804-2	Sequence 2, Appl
28	165	5.0	724	15	US-10-023-634-80	Sequence 80, Appl
29	164	5.0	1790	15	US-10-369-493-1586	Sequence 1586, Appl
30	163.5	4.9	308	12	US-10-425-114-64137	Sequence 64137, A
31	163.5	4.9	498	15	US-10-104-047-3191	Sequence 3191, Appl
32	163	4.9	287	14	US-10-029-386-33041	Sequence 33041, A
33	162.5	4.9	709	14	US-10-025-250-15	Sequence 15, Appl
34	162.5	4.9	709	15	US-10-023-634-14	Sequence 14, Appl
35	162	4.9	1992	15	US-10-369-493-6527	Sequence 6527, Appl
36	161.5	4.9	621	15	US-10-108-260A-4409	Sequence 4409, Appl
37	161.5	4.9	1875	15	US-10-369-493-22285	Sequence 22285, A
38	161.5	4.9	3878	14	US-10-080-608A-11	Sequence 11, Appl
39	161.5	4.9	3907	14	US-10-171-311-2	Sequence 2, Appl
40	161.5	4.9	3911	15	US-10-370-685-100	Sequence 100, Appl
41	161.5	4.9	3925	14	US-10-171-311-6	Sequence 6, Appl
42	161	4.9	650	15	US-10-104-047-3636	Sequence 3636, Appl
43	161	4.9	725	10	US-09-978-309A-47	Sequence 47, Appl
44	161	4.9	1837	15	US-10-369-493-22734	Sequence 22734, A
45	160.5	4.8	621	14	US-10-316-253-44	Sequence 44, Appl

ALIGNMENTS

```
RESULT 1
US-10-425-114-58390
; Sequence 58390, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58390
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB3079-030-A8_FLI.pep
; US-10-425-114-58390

Query Match 43.1%; Score 1428.5; DB 12; Length 611;
Best Local Similarity 47.2%; Pred. No. 2.4e-99;
Matches 294; Conservative 99; Mismatches 159; Indels 71; Gaps 15;

QY 33 RLASQDDGGE-----WEVISKKNKPKNTSGKTW-----VSQNSPPRA 73
DB 13 RSARSWSGSGDRRGCGPSPSHSGMETWKKSK-KPQAGAGQWAPWSSTVTENTARP-A 70
QY 74 WGGQOQGRGSNYS---GRANNVSGRNGRGRIQANISGRGALSRYKYNFPAP----- 126
DB 71 WGGSGSHSPSGTSGWAQPDHGAATRNPR-----PPSQTS 105
QY 127 -PVSRRPLSGMWMQARGSGAGHTAVOE-----PPVED-DVDNASEENDSDALDSD 177
```

```

Db      106 RPLAFLPANGWQMRPRPSGSEVKKDDAPPSGVSPEVENGNNTSDDDDD--DDL 163
Qy      178 DDDLADDDVDSVQSHSGSRKKNKPKPRGSLDLSLSTQINEPQRMHCPACQNGPRA 237
Db      164 SDDI--SDDVDSDAEKSFEFRKTNKWFKEFEVLNTLSLEQINEQRMHCPACKNPG 222
Qy      238 IDMY-NLHPLLAHARTKGARVYKLHRELAEVLKEDLOMGASVTPCGEYIGQKGLGEDE 296
Db      223 IDMYKLOPLVSHARKTRVYKLHRELAALBEEISRRTSTVLPAGEOPGKKGL-QDS 281
Qy      297 KQYEIWPMPVVIIMNTRLDKDDNDKWLGMGNOLLEYPDKYEALRAHSHYSGPGRHGMV 356
Db      282 TBREIWPMPVVIIMNTRFLEKEDDDKWKGMGNOLLDYFGEYEASKARHAYGSGRHGMV 341
Qy      357 LMFESSATGYLEAERLHRELAEMGLDIRIAMGQKRSNP--SGGVOLYGFATKDDLDLFPQ 415
Db      342 LTFESSAVGMEAEERLHKFVNOGTDRNSWHLRKVFPVPGKQLYGFLANKEDMEAFNK 401
Qy      416 HSQKTRLKFEKLSYQBMVVKELROI SEDNOQLNFPKNLSKONKAKVLEESLEIMSEK 475
Db      402 HCHGSRKLTKEKMSYEMVVIQKMS SEDNOQLNFKKNVKTQKSKAVESLGVYTQK 461
Qy      476 LRRTAEDNRIVRQRTQMOHEONREENDAHDRFPMDSIKOIHERRDAKENFEMLQOQERA 535
Db      462 LRTEIENIFVRSKAKKHEVEEMKSGOEI FPHGLIEDIHKATEDKQFELQLQERS 521
Qy      536 KVVGOQOQINPSSNDCKRAEVSFIEFOEKMEEFVEEREMLIKQOKEKMEEMKKR 595
Db      522 KA---RRFVDSGTMDRQLRKRYVQKFI DCQVDAEVEFVDELIVKHEDKRLKKE 578
Qy      596 HHEIFDLEKEFDEALEQMLYKH 618
Db      579 YMDLELEKEFPAALTGLMEKH 601

```

```

RESULT 2
US-10-424-599-234018
; Sequence 234018, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234018
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1.pcp
US-10-424-599-234018

```

```

Query Match      22.5%; Score 745.5; DB 12; Length 324;
Best Local Similarity 31.1%; Pred. No. 4.3e-48;
Matches 162; Conservative 68; Mismatches 84; Indels 207; Gaps 6;

Qy      109 GGRALSRKYDNNFVAPPPVSRPPLGGMNMQARGSAQHTAVQER--PDVEDD--VDNA 164
Db      1 GGPQLTSGESNYVTNPNVIRPPLRHGMNMQSRPAIOSNVDRDISPEDLOKNYGVDD 60
Qy      165 SEENDSDALDDDDDLASDDVDVDSVQSHSGSRKKNKPKPRGSLDLSLSTQINEPQRMH 224
Db      61 GEEB--ESDLEDTDDDDDDSDASQSHSTRKSKKKYKFFELLDGLTVEQINEPER 119
Qy      225 QNHCPACQNGPAIDMY-NLHPLLAHARTKGARVYKLHRELAEVLKEDLOMGASVTPCG 283

```

```

Db      120 QNHCPACQNGPAIDMYNRLQPLVTHAKTGSKRYKIHRELAILLDEELRRRGSVTPPA 179
Qy      284 EIVGQKGLGEDEKQYEIWPMPVVIIMNTRLDKDDNDKWLGMGNOLLEYPDKYEALRAH 343
Db      180 EVFGKWNGL----- 188
Qy      344 HSYGQGHGMVLMFESSATGYLEAERLHRELAEMGLDIRIAMGQKRSNMFSGGVOLYGF 403
Db      189 ----- 188
Qy      404 LATKDDLDIFNOSHQKTRLKFEKLSYQBMVVKELROI SEDNOQLNFPKNLSKONKAK 463
Db      189 ----- 188
Qy      464 VLEESLEIMSEKLRRTAEDNRIVRQRTQMOHEONREENDAHDRFPMDSIKOIHERRDAKE 523
Db      189 -----KVQOQENKEEMSMOEFPFEQIRTIHDSAAKE 221
Qy      524 ENFEMLOQOERAKVVGQOQOINPSSNDCKRAEVSFIEFOEKMEEFVEEREMLIK 583
Db      222 EEFERNQOQEKREK---KRSSTSPINEBGRKYVDYLKFEVFOQKEMENFVAEBEKLRQ 278
Qy      584 DQEKMEEMKKSHHEIFDLEKEFDEALEQMLYKHGLHNE 624
Db      279 AHKDVNDAMTRRRHWEKVOLEERFNEELAKLMEKXSLSHPE 319

```

```

RESULT 3
US-10-424-599-234019
; Sequence 234019, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234019
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(223)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1.pcp
US-10-424-599-234019

```

```

Query Match      21.8%; Score 721; DB 12; Length 223;
Best Local Similarity 60.1%; Pred. No. 1.9e-46;
Matches 131; Conservative 41; Mismatches 40; Indels 6; Gaps 4;

Qy      109 GGRALSRKYDNNFVAPPPVSRPPLGGMNMQARGSAQHTAVQER--PDVEDD--VDNA 164
Db      1 GGPQLTSGESNYVTNPNVIRPPLRHGMNMQSRPAIOSNVDRDISPEDLOKNYGVDD 60
Qy      165 SEENDSDALDDDDDLASDDVDVDSVQSHSGSRKKNKPKPRGSLDLSLSTQINEPQRMH 224
Db      61 GEEB--ESDLEDTDDDDDDSDASQSHSTRKSKKKYKFFELLDGLTVEQINEPER 119
Qy      225 QNHCPACQNGPAIDMY-NLHPLLAHARTKGARVYKLHRELAEVLKEDLOMGASVTPCG 283
Db      120 QNHCPACQNGPAIDMYNRLQPLVTHAKTGSKRYKIHRELAILLDEELRRRGSVTPPA 179
Qy      284 EIVGQKGLGEDEKQYEIWPMPVVIIMNTRLDKDDNDK 321
Db      180 EVFGKWNGLGEDEKQYEIWPMPVVIIMNTRLDKDDNDK 217

```



```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63998
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3245-195-E3_FLI.pep
; US-10-425-114-63998

```

```

Query Match      6.0%; Score 199.5; DB 12; Length 414;
Best Local Similarity 20.8%; Pred. No. 1.1e-06;
Matches 97; Conservative 97; Mismatches 185; Indels 95; Gaps 22;

```

```

Db      142  RCGSAOHTAVQEPFVEDVDVNAEEENDSDALDDSDDLASDDVDVDSQKSHGRKON 201
      12  RASGHQVAYFFPFLK--MDHSDSDTE--ISDSE--IDYEG----- 49
      202  KWFKKFPGSLDLSIBOINERPOHCPACONGFGAIDWYNHPLLAHR-----TKGA 255
      50  ---KIVARLMSGDL-KLNKNGENYICPFCSKKKN---YKSSILVQHAGVSAAPENREA 101
      256  RRVKLRHRLAELVLEKDLQMR---GASVTPCGEIVGQKGLGDEKDEYIWPMPVIMNT 312
      102  KKVAVHRLAFKYLKNGKLAKSPESGSHATPV-----EPOLQNRDKFVPMNGVIVNV 154
      313  RLDKDNNDKVLGNGQELLEYFDKYEALRAHSYGPOGHRMSVYLMESSATGYLEAERL 372
      155  PTEWKDG-RQIESGNRLKEQLSHCPARVPIWTFRGHTGNAIVFEGKWMNGFRNARTP 213
      372  HRELAEMGDIRIWMQKSMFSGVROLYGLATKQD---INQHSQKTRLKPELKS 429
      214  ESHFPAAGFGKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 265
      430  YQEMVYKELRQISENQOQLNFKYKSKONKAKYLESELSIMSEKLRRTAEDNRIVQR 489
      266  VNDLAEGAR---KNDRLVANIANOJEVKNRYLQELF-----CYSETTSLERKMGOR 316
      490  TKMQHQRNEEMDANDRPFWDISIKQIHERDAK--EENFEMLOQOERKAVVGGQOQNINP 547
      317  EQLLQSTNEB-----ISKWQQLARRHSQKIIDEN-QKLRSDLEAKM----- 356
      548  SSND-DCR-KRAEVSFIFFOKEMEFEVEEREMLIKDOEKKMDKKRHHEE 599
      357  --NDLDVRSKQDEIAKASDYSRRSLQ--EKQKNAIKSHLKLNTLEQKXADE 406

```

```

RESULT 8
US-10-425-114-54044
; Sequence 54044, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54044
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Zea mays

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17282F05_FLI.pep
; US-10-425-114-54044

```

```

Query Match      6.0%; Score 199.5; DB 12; Length 448;
Best Local Similarity 27.0%; Pred. No. 1.2e-06;
Matches 62; Conservative 40; Mismatches 81; Indels 47; Gaps 8;

```

```

Db      228  CPACONGPGALDWNHPLLAHARTG-----ARVYKLRHRLAELVLEKDL-----QMRGA 277
      224  CLAC--GSSKDPADVHGLVMEVAVPNADSLVHGLHRLAELVLEKDL-----QMRGA 281
      278  SVTPCGEIVGQKGLGDEKDEYIWPMPVIMNTRLDKDNKVLGNGQELLEYFDK 336
      282  QSLP-----SDLVQASREDLIWPPTVLIHMTSTRRKQDGLBGLGNEM---DKK 329
      337  ---YALRAHSYGPOGHRMSVYLMESSATGYLEAERLHRELAEMGDIRIAMQ--- 388
      330  MTELFGSGKSKLYGKEGMGLTLIKFANNSSGLKEARLAEYLERODRGRIGMSRAHA 389
      389  KRSMFGS-----GVROLYGLATKQDLDIFNHSQKTRLK 424
      390  SRVSDQNPDLVETRTAERKRLIYGLALASDLDELSDSRKASLIQ 439

```

```

RESULT 9
US-09-839-479-21
; Sequence 21, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-21

```

```

Query Match      5.5%; Score 182; DB 12; Length 1972;
Best Local Similarity 20.8%; Pred. No. 0.00018;
Matches 113; Conservative 83; Mismatches 190; Indels 158; Gaps 20;

```

```

Db      140  QABGSAOHTAVQEPFVEDVDVNAEEENDSDALDDSDDLAS----- 183
      394  QFRCTSDITSSKSDSDNDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 453
      184  --DDYSDVSQKSHGRKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 239
      454  DDDDKQDESDSDTEGKTSKMLNTTSSVKSQPSMSTGHSFPRNHLIAKAGSAPALIC 513
      240  WYNHPLL-----AHARTKGARVYKLRHRLAELVLEKDL-----LOWR 275
      514  SESQPAFLGTSSSTLSSHSQTSKRRRYTDERELRIPLEYEGWORETRIRNPGRLQGE 573
      276  GASVTPCGEIVGQKGLGDEKDEK-----KGLGDEKDE-----YEIVW--- 303
      574  VAYYAPCGKRLRQYDEVYKYLNRNGIMDISRDNFSGSAKTRVDYFLEADGPQEMQCLL 633
      304  ---PPTVIMNTRLDNDKDKVLGNGQELLEYFDKY---DALRAHSYGPOGHRMS 355

```

Db 634 KEDVPIRIRAMEGRGRPPNP-----DROARRESESMRRRKGPRPNVNA 679
 QY 356 VLMFESSA--TGYLEARLHRELAEMGLDIRIAMGOKRSMPSGGVROLYGLATKODLDIF 413
 Db 680 EFLDNDADLKLRLKQOEIARQAQIKLRLKLOKQOARVAKAKQOALMAEE----- 734
 QY 414 NOHSQKTELKPELK-SYQEMVVKELROI SEDN---QOLNFKNLSKONKRAVLEES 468
 Db 735 -----KRQKQKIKIMKQOEKIKRIQOI RMEKELAAQOILEAKKKKKKEBAANAAXILLEAE 788
 QY 469 LEIMSEKLRTLEDNRIVQRTYMOHEONREEMDAHDFRPMOSIKOIHERRDAKEENFEM 528
 Db 789 KRIKKEKEMER-----QOAVLLKH-QERERRQH-----MMLKAMEARKKAEK--ER 833
 QY 529 LQOGEAKVVGQOQONINSSNDCKRAEVSFIIEFOKEMEVEEVEEMLIKQOEK 588
 Db 834 LKOEK-----DEKRLN-----KERKLEORLEEM-AKELKXP 866
 QY 589 MEDM 592
 Db 867 NEDM 870

RESULT 10
 US-10-097-340-43
 ; Sequence 43, Application US/10097340
 ; Publication No. US20030087250A1
 ; GENERAL INFORMATION:
 ; APPLICANT: John MONAHAN
 ; APPLICANT: Manjula GANNAVARAPU
 ; APPLICANT: Sebastian HOERSTCH
 ; APPLICANT: Shubhangi KAMATKAR
 ; APPLICANT: Steve G. KOVATS
 ; APPLICANT: Rachel E. MEYERS
 ; APPLICANT: Michael MORRISSEY
 ; APPLICANT: Aml SEN
 ; APPLICANT: Peter OLANDT
 ; APPLICANT: Peter VEIBY
 ; APPLICANT: Gordon B. MILLS
 ; APPLICANT: Robert C. BAST, JR.
 ; APPLICANT: Karen LU
 ; APPLICANT: Rosemarie SCHMANDT
 ; APPLICANT: Xumei ZHAO
 ; APPLICANT: Karen GLATT
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
 ; FILE REFERENCE: MRI-030
 ; CURRENT APPLICATION NUMBER: US/10/097,340
 ; PRIOR FILING DATE: 2002-03-14
 ; PRIOR APPLICATION NUMBER: 60/276,025
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/325,149
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/276,026
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/324,967
 ; PRIOR FILING DATE: 2001/09/26
 ; PRIOR APPLICATION NUMBER: 60/311,732
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/325,102
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/323,580
 ; PRIOR FILING DATE: 2001-09-19
 ; NUMBER OF SEQ ID NOS: 363
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 43
 ; LENGTH: 1203
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-340-43
 Query Match 5.4%; Score 178; DB 14; Length 1203;
 Best Local Similarity 19.0%; Pred. No. 0.00019;

Matches 173; Conservative 121; Mismatches 284; Indels 332; Gaps 40;
 QY 15 OGQYRPE-----VEOLVOGLGTR--LASODDGGEMEYISKONKNGKSGTKM 63
 Db 37 RGRPRPAKARASTYGVAVRVGIGIAQPPVVLNSGEGSDSGVOIKGANDG-ASGA-- 93
 QY 64 VSQNSNPAPWAGQOQOGRGSNVSGRNNVSGRNGNGRGI-----QANISGRGALSRYKD 119
 Db 94 LSSDLELPNPYSQVKGFPAPQSOSTSDEEPGAYNNKGLRSHSQASLAGPQVDSNRS 153
 QY 120 NNFV-APPVSRP-----PLE-----GKNNQARGGSAOHTVAQEPDYVE 158
 Db 154 NSMTLEIAPVYASPGSTIDTAPLSSVDSLINKPDSQIG--QARGTRGRTRM--LPPEQ 208
 QY 159 DDVDNASEENSDALDDDDDD-----LASDDYDSVSGKSHGSRKONKMPKFFSLS 213
 Db 209 RRRKSLDSRLPDDTFFEREROSTNHTWTSSTYKDNHVTGSKOPASQON-----LSP 259
 QY 214 LS-----IEQINEPOROMHCPACON-----GPAIDWY--NLH 244
 Db 260 LSGFSRSRQTQDWVLQSFEPFRSAODPTMLQFKSTPDLRRDQGEAAPGGSVDHMKATY 319
 QY 245 PLIAHARTKGAARRVKILHRELAYLEKDLQMRGASVTPGE--ITYQ-----WIG 291
 Db 320 GILREGSSESETSVR--RKVSIVLEK--MQLVWVSSGSTAVAGGELTRKVEBLQK 374
 QY 292 LGEDKDYEIWPPWYIINNTBLDKDNDKWLGMGNQELLE-----Y 333
 Db 375 LDEEVKKRKLKPSQVGLERQLEKTEBCSRL--QELERRKGAQOSKELQNMKEL 430
 QY 334 PDYEALR-----ARHSYQO----- 349
 Db 431 LQGGEDLRHGLTQWELQNKLKHVGPRPAKEVLLKDLLETRELLLEVLEBQRYVEQL 490
 QY 350 --GHRGSVYL--MFESSAT-----GYLEARR-- 371
 Db 491 RLRERELTLALKALKEEVASRDQEVHVRQYQYORDTEQRRSMQDATODHAVALERQKM 550
 QY 372 -----LHRELAEMGLDIRIAMGOKRSMPSGGVROLYGLATKODL----- 410
 Db 551 SALVRLGLQRELETSSETGHW--OSMFOKNKEDL--RATQOELLQRLMEKEMEELG 604
 QY 411 ---DIFNOH-----SOGKTR--LKFEKSYQEVVVKEL----- 438
 Db 605 EKIEVQLRELEQARASAGTROYEVYLKELLTQF-ELKELOAERQSOVAGHRHRELE 663
 QY 439 -----ROISEDNOQ--LNYFNKLSKONKRAVLEESLIMSEK----- 475
 Db 664 KOLAVLRYEADRGRELEBQNLQLOKTLQRLQDCEBASAKVVALEAATVLCGRRAAVET 723
 QY 476 -LRTFEDNRIRYQRTYMOHEONREEM-----DAHDRFPMOSIKOIHERRDAKEENFEM 528
 Db 724 TLRETOEENDEFFRRRLGLEQQLKETRGVLVDGEAVALRLKLOLEAKEQOLEBALNA 783
 QY 529 LQOGE-----RAKV-----VGQOQONINPSNDDCKRAEVSFIIEFOR- 568
 Db 784 SQEBEGSLAAARALEARLEBAQGLARLQGOQTLNRLYBEGKREVLRLKQALEBQ 843
 QY 569 -----KEMEEFVEEREMLIKQOEKMEKMKHHEIFPLEKEF-DEALEQIM 615
 Db 844 KRLDRTVRLNKKLEKIGEDSKQALQOQALQLEDEYKFAREVDACQARDWASEAK 903
 QY 616 YKHGLHNEED 625
 Db 904 TSGGLSRLQD 913

RESULT 11
 US-10-296-115-1265
 ; Sequence 1265, Application US/10296115
 ; Publication No. US20040053248A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq Inc

TITLE OF INVENTION: No. US20040053248a1el Nucleic Acids and Polypeptides
 FILE REFERENCE: 784PCT
 CURRENT APPLICATION NUMBER: US/10/296,115
 CURRENT FILING DATE: 2002-11-18
 PRIOR APPLICATION NUMBER: US09/488,725
 PRIOR FILING DATE: 2000-01-21
 PRIOR APPLICATION NUMBER: US09/552,317
 PRIOR FILING DATE: 2000-04-25
 NUMBER OF SEQ ID NOS: 1478
 SEQ ID NO 1265
 LENGTH: 1879
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)..(1879)
 OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3
 US-10-296-115-1265

Query Match 5.3%; Score 175; DB 12; Length 1879;

Best Local Similarity 20.2%; Pred. No. 0.00057;
 Matches 116; Conservative 94; Mismatches 194; Indels 170; Gaps 24;

158 EDDVNNASEE-----ENDSDALDDSDDDLASDDYSDVSQKSHGRKONKWFK 206
 Db 1297 QKEIENLTQYBEKAAAYDKLEKTKNRLQDELDDLIV-DIDNOROLVSNLEKQKQ-FDQ 1354
 QY 207 FFGSLDSLSIEQINEPOROMHCPCACONGPAGALDWNHPLLAARTGAGRVKLHRELAE 266
 Db 1355 LLAERKNISSKYABERDR-----VEAEKREKETALDLALAE 1393
 QY 267 VLEKLOMRGASVIRPCGEI-----YQWKGLESD-EKDYEIWPPMVIINTRLD 315
 Db 1394 ALEAEVELEKRYMKLAKEMRGASAKDVGQ-ELSHLEKSKRALGDPLEEMKQLE 1451
 QY 316 KDDNDKMLGMNQEL-----LEY-----PDKYEALRAHSGVQPGHGMV 356
 Db 1452 E-----LGRTELASPRDAKRLLEVNMQAPSPASPER--DLOARTQONEERHLQR 1501
 QY 357 LMFESATGYLAERLHRELAEMLGDLRIAMGQRSM-----FSGG--VROLYGFIA 405
 Db 1502 QLHEVETE--LEDERKORALAAAKIKLGMDPVRTLDXADSAIKRGKAKIKQLRLKLOA 1559
 QY 406 TKODL-----DIF--NOSHOGKTLKELKSYQWVYKELRQISENQOLNY 450
 Db 1560 QMMDPQRELEDAFASDEIFATAKENKAKSLIADLQLOE-----DLAAEBGRKQADL 1615
 QY 451 FKXKLS-----KONHAKY--LEESL-----EINSEKLAR 478
 Db 1616 EKELEAEELASSISGRNALODEKRLERARLQLEELLEEQNMESMDRVAKATQQAEO 1675
 QY 479 -----TAEDNRLVRQRTKQHEQNEEMDADRPFMDSIKQIHERRAK-----EE 524
 Db 1676 LSNELATERSTAOKNESARQOOLERONKELSKLHEWGAVKSKFTTIALDAKLAQLE 1735
 QY 525 NFEMLQOERAKVYGOQONINPSSNDCKRAEYVSSFIEQEKMEEFVERE----- 579
 Db 1736 QVEQEAERKQATSLKQKD-----KALKETILQVEDBRKQAEQYKEQAKGNAR 1785
 QY 580 -MLIKDOEKMEKMKRHEEIPDLKEFDEALE 612
 Db 1786 VKQLKQLEAEESQSRINARRLQRELEDEATE 1819

RESULT 12

US-09-839-479-71
 Sequence 71, Application US/09839479
 Publication No. US20020039779A1
 GENERAL INFORMATION:
 APPLICANT: Jones, Michael H.
 TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
 FILE REFERENCE: 06501-042002
 CURRENT APPLICATION NUMBER: US/09/839,479

CURRENT FILING DATE: 2001-04-20
 PRIOR APPLICATION NUMBER: US 09/418,710
 PRIOR FILING DATE: 1999-10-15
 PRIOR APPLICATION NUMBER: PCT/JP98/01783
 PRIOR FILING DATE: 1998-04-17
 PRIOR APPLICATION NUMBER: JP 9/310027
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: JP 9/116570
 PRIOR FILING DATE: 1997-04-18
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 71
 LENGTH: 1969
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-839-479-71

Query Match 5.2%; Score 173.5; DB 12; Length 1969;

Best Local Similarity 20.1%; Pred. No. 0.00079;
 Matches 128; Conservative 94; Mismatches 213; Indels 203; Gaps 25;

47 ISKKNKKPNTSGKTWVSONSPRAWGQOQGRGSNVSGKNNVSGKNGC----- 100
 Db 341 LSTSGRRTPGN---QTPWPSAPIL---HSQGEKAVSNVNVVKTQHSHPKSLVE 393
 QY 101 --RGIQNISGRGRALSRKYDNFVAPPVSRPPLGGMNMGARGSAGHTAVQFPVVE 158
 Db 394 QFRGTSDI-----PSSK-----DSEDSNDEEDDEE 421
 QY 159 DDVNNASEENDSDALDS-----DDDLASDDYSDVSQKSHGRKONKWFK 205
 Db 422 EDEBEDDESDDSESDDSSSESDTGESBEDDDDDQDESDSD--TEGETSKYKMK 477
 QY 206 KFGSLDSLSIEQINEPOROMHCPCACONGPAGALDWNHPL-----AHARTG 254
 Db 478 TTSSKSPMSLTGHTSTFNLHIAKAPGAPALCSEQSPALGTSSSTLTPSPHGCTK 537
 QY 255 ARVYKLHRELAEVLEKD-----LQNRGASVIRPCGEIYQW-----KGL 292
 Db 538 RRRVYDERELRLPLEYGMQRETRIRNFGRLQGSVAIYVACGKKLRQYPPVIRKYLIRNGI 597
 QY 293 GEDEKD-----YEIVW-----PPMVIINTRLDXDNDKMLG 324
 Db 598 MDSRNPFSKAKIRVGDFFEARDPQEMQWCLLKEDVYPRIRAMGRGRGRPNP----- 653
 QY 325 MGNQELLEYFDKY---EALPARHSGPQGHGMSVLMPESSA--TGYLEAERLHRELAE 379
 Db 654 -----DRORAEESRMRKRRKRPVGNNAFLDNADAKLRLKLOQEA-ROAAOI 702
 QY 380 GLDRIAMGQRSMFSGVRYQLYFLATKODDIFNOSHOGKTLKELKSY-QEMVYKEL 438
 Db 703 KLRKLOKQOQARVAKAEAKQQAIVAAE-----KRKOEQIKHKQOQKIKRI 751
 QY 439 ROISEDN---QOLNYFKKLSKONKHAUVLEESLEIMSEKLRTAEDNRLVRQRTKQ 494
 Db 752 QQIRMEKELRAQOQILAKKKKEEANAALDAEKRIKEREK-----RQAAVLK 802
 QY 495 EQREEMDADRPFMDSIKQIHERDAKENFEMLOQOERAKVYGOQONINPSSNDCK 554
 Db 803 ROERERRRQH---MMLKMGEMARKKAEEK--ERLQOERK-----DEKRLN----- 842
 QY 555 KRAEYVSSFIEQEKMEEFVEREMLIKDOEKMEEM 592
 Db 843 -----KERKLEQRRLLEEM-AVELKPNEDM 867

RESULT 13

US-10-363-616-295
 Sequence 295, Application US/10363616
 Publication No. US20040044181A1
 GENERAL INFORMATION:
 APPLICANT: Hyseq, Inc
 TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

3:	gd_in:
4:	gd_om:
5:	gb_ov:
6:	gb_pat:
7:	gb_ph:
8:	gb_pl:
9:	gb_pr:
10:	gb_ro:
11:	gb_srs:
12:	gb_sy:
13:	gd_un:
14:	gb_va:
15:	em_ba:
16:	em_fun:
17:	em_hum:
18:	em_in:
19:	em_mu:
20:	em_or:
21:	em_ov:
22:	em_ov:
23:	em_pat:
24:	em_ph:
25:	em_pl:
26:	em_ro:
27:	em_srs:
28:	em_un:
29:	em_vi:
30:	em_hlg_hum:
31:	em_hlg_inv:
32:	em_hlg_mus:
33:	em_hlg_other:
34:	em_hlg_pln:
35:	em_hlg_rtd:
36:	em_hlg_mam:
37:	em_hlg_vrt:
38:	em_sy:
39:	em_hgo_hum:
40:	em_hgo_mus:
41:	em_hgo_other:

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	3275	100.0	3275	6	AX078760	Sequence
	2	3275	100.0	81365	8	AB025633	Arabidops
	3	2254	69.8	2254	8	AF239719	Arabidops
	4	972	29.7	2162	8	BT002944	Arabidops
	5	963	29.4	1878	6	AX078761	Sequence
	6	963	29.4	1909	8	BT004380	Arabidops
	7	541	16.5	650	8	ATH528171	Arabidops
	8	45	45	45	8	ATH527741	Arabidops
	9	24	0.7	170765	2	AC102284	Mus muscu
	10	23	0.7	554	3	AF057056	Uroloeucon
	11	23	0.7	101080	2	BX0004860	Danio rer
	12	23	0.7	101241	5	AL845510	Zebrafish
	13	23	0.7	126716	9	AC107028	Homo sapi
	14	23	0.7	152345	5	BX005128	Zebrafish
	15	23	0.7	180925	2	AC048381	Homo sapi
	16	22	0.7	2100	3	D84313	Drosophila
	17	22	0.7	6136	6	AX348482	Sequence
	18	22	0.7	6301	6	AX344692	Sequence
	19	22	0.7	9888	6	AX346143	Sequence
	20	22	0.7	80318	5	AL645798	Zebrafish
	21	22	0.7	81835	8	ATFC9	Arabidops
	22	22	0.7	110000	2	AC107170	1- Continuation (2 of
	23	22	0.7	131020	9	HS393B23	Human DNA s
	24	22	0.7	141509	9	AC006044	Homo sapi
	25	22	0.7	150030	2	AC145578	Mus muscu
	26	22	0.7	154923	3	AC026185	Homo sapi
	27	22	0.7	158661	10	AC074229	Mus muscu
	28	22	0.7	174612	9	AC018494	Homo sapi
	29	22	0.7	174641	2	AC016346	Homo sapi
	30	22	0.7	175569	10	AC109214	Mus muscu
	31	22	0.7	177529	2	AC119271	Mus muscu
	32	22	0.7	177834	2	AC018811	Homo sapi
	33	22	0.7	178928	2	AC121725	Rattus no
	34	22	0.7	179134	9	AC119607	Rattus no
	35	22	0.7	181166	9	AC087897	Homo sapi
	36	22	0.7	183509	9	AC112211	Homo sapi
	37	22	0.7	183936	10	AC007978	Mus muscu
	38	22	0.7	191391	2	AC128938	Rattus no
	39	22	0.7	192173	9	AC107462	Homo sapi
	40	22	0.7	193523	5	BX001014	Zebrafish
	41	22	0.7	197909	10	AL928719	Mouse DNA
	42	22	0.7	198067	8	ATCHRIV47	Arabidops
	43	22	0.7	201208	2	AC107451	Mus muscu
	44	22	0.7	201803	2	AC113572	Canis fam
	45	22	0.7	205209	2	AC137942	Mus muscu

ALIGNMENTS

RESULT 1					
AX078760	AX078760	3375 bp	DNA	linear	PAT 22-FEB-2001
LOCUS	AX078760				
DEFINITION	Sequence 1 from Patent WO0105951.				
ACCESSION	AX078760				
VERSION	AX078760.1 GI:13158379				
KEYWORDS	.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Euastrata; Viridiplantae; Magnoliophyta; eudicotyledons; Core eudicots;				
	Spermatophytes; Magnoliophyta; Brassicales; Brassicaceae; Arabidopsi.				
	s. sensu stricto; Arabidopsis				
REFERENCE	Beclin,C., Elmayer,T. and Vaucheret,H.				
AUTHORS	Novel sgas plant gene and use thereof				
TITLE					

JOURNAL Patent: WO 0105951-A 1 25-JAN-2001;
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)

FEATURES
 source Location/Qualifiers
 1.3275
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"

primer_bind
 primer_bind
 complement(2926..2952)
 /note="p356d"
 /note="p356f"

ORIGIN

Query Match 100.0%; Score 3275; DB 6; Length 3275;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GACAAACAAACAAATTAAGCAAGTCATGTCGTGCAATTAATTAATGAGGAAACA 60
 1 GACAAACAAACAAATTAAGCAAGTCATGTCGTGCAATTAATTAATGAGGAAACA 60
 61 TTAAGTTAAGCGGAAAGGAAAGGTAACAAATGAAACAAATCAACTGAAT 120
 61 TTAAGTTAAGCGGAAAGGAAAGGTAACAAATGAAACAAATCAACTGAAT 120
 61 TTAAGTTAAGCGGAAAGGAAAGGTAACAAATGAAACAAATCAACTGAAT 120
 121 GAAATTTGAGTCAGATCGAATCGAAGGCGGTTTAAAGCTTAATTAAGCTTCTC 180
 121 GAAATTTGAGTCAGATCGAATCGAAGGCGGTTTAAAGCTTAATTAAGCTTCTC 180
 121 GAAATTTGAGTCAGATCGAATCGAAGGCGGTTTAAAGCTTAATTAAGCTTCTC 180
 181 ATTGTCCTCTTCTGTCAGTTATTTCTTCTCCGGAATCCTGACTACATCTCTAC 240
 181 ATTGTCCTCTTCTGTCAGTTATTTCTTCTCCGGAATCCTGACTACATCTCTAC 240
 181 ATTGTCCTCTTCTGTCAGTTATTTCTTCTCCGGAATCCTGACTACATCTCTAC 240
 241 TCTCCGCGGCTTAACTTAAGTCTCTCCGCTTAACTTAAGTCTCTCCGCTTAA 300
 241 TCTCCGCGGCTTAACTTAAGTCTCTCCGCTTAACTTAAGTCTCTCCGCTTAA 300
 241 TCTCCGCGGCTTAACTTAAGTCTCTCCGCTTAACTTAAGTCTCTCCGCTTAA 300
 301 CCTCCGATGCGCTCAGCGCATGATCTGTCGATTTCTTTTCTTCTGCTGAAAA 360
 301 CCTCCGATGCGCTCAGCGCATGATCTGTCGATTTCTTTTCTTCTGCTGAAAA 360
 301 CCTCCGATGCGCTCAGCGCATGATCTGTCGATTTCTTTTCTTCTGCTGAAAA 360
 361 ATTGCTTAATGTCGATTTGCAAGGTTTGTGCTAAGGTTACTTTTCCCTATA 420
 361 ATTGCTTAATGTCGATTTGCAAGGTTTGTGCTAAGGTTACTTTTCCCTATA 420
 421 TTTATAGTCTTGAATACGATCTGCTCTTACTGTTTGTGCTAATTTGTTGCT 480
 421 TTTATAGTCTTGAATACGATCTGCTCTTACTGTTTGTGCTAATTTGTTGCT 480
 421 TTTATAGTCTTGAATACGATCTGCTCTTACTGTTTGTGCTAATTTGTTGCT 480
 481 TTCAACGTTTATGTCGATCGAGTATTTGACTGTGAAAAATCCCTCGTTTGGT 540
 481 TTCAACGTTTATGTCGATCGAGTATTTGACTGTGAAAAATCCCTCGTTTGGT 540
 481 TTCAACGTTTATGTCGATCGAGTATTTGACTGTGAAAAATCCCTCGTTTGGT 540
 541 TTTGTTTCAATTAATGCAATGATCTACCTTTTGCTTGAATTTGTTTGAAGCT 600
 541 TTTGTTTCAATTAATGCAATGATCTACCTTTTGCTTGAATTTGTTTGAAGCT 600
 541 TTTGTTTCAATTAATGCAATGATCTACCTTTTGCTTGAATTTGTTTGAAGCT 600
 601 ATGCTTGTGCTGTTGTTATTAATCACTTCACTGATGATTTGAGATTTGATG 660
 601 ATGCTTGTGCTGTTGTTATTAATCACTTCACTGATGATTTGAGATTTGATG 660
 601 ATGCTTGTGCTGTTGTTATTAATCACTTCACTGATGATTTGAGATTTGATG 660
 661 ACTGTTGCTTCTTGTGCTATATGTTGTTAAAAATGAGTTCTAAGGCTGCTCA 720
 661 ACTGTTGCTTCTTGTGCTATATGTTGTTAAAAATGAGTTCTAAGGCTGCTCA 720
 661 ACTGTTGCTTCTTGTGCTATATGTTGTTAAAAATGAGTTCTAAGGCTGCTCA 720
 721 CTAAGGAAAGAAAGCTTCAAGGCTGTTAAGGCTGAGGTTGAACAGTTGCTCA 780
 721 CTAAGGAAAGAAAGCTTCAAGGCTGTTAAGGCTGAGGTTGAACAGTTGCTCA 780
 721 CTAAGGAAAGAAAGCTTCAAGGCTGTTAAGGCTGAGGTTGAACAGTTGCTCA 780
 781 TGGCAGGAGCAAGAGCTGCTTCTTCAAGATGATGAGAGAGGAGGCTCATTTCCA 840
 781 TGGCAGGAGCAAGAGCTGCTTCTTCAAGATGATGAGAGAGGAGGCTCATTTCCA 840

841 AGAAGAACAAAGAACAAACGAGAAACACTTCTGAAAAAATCTGGGTTCTCAGATTCGA 900
 841 AGAAGAACAAAGAACAAACGAGAAACACTTCTGAAAAAATCTGGGTTCTCAGATTCGA 900
 901 ATCTCTAGAGCTTGGGTTGTCAGCAGCAGAGGAGGATGCAACGATCTGGAGAG 960
 901 ATCTCTAGAGCTTGGGTTGTCAGCAGCAGAGGAGGATGCAACGATCTGGAGAG 960
 901 ATCTCTAGAGCTTGGGTTGTCAGCAGCAGAGGAGGATGCAACGATCTGGAGAG 960
 961 GAAACAAATGATCCGAGAGGATGCAACGATGTCGGGCTTCAAGCTATATCTG 1020
 961 GAAACAAATGATCCGAGAGGATGCAACGATGTCGGGCTTCAAGCTATATCTG 1020
 961 GAAACAAATGATCCGAGAGGATGCAACGATGTCGGGCTTCAAGCTATATCTG 1020
 1021 GTGCGGAGCAGGCTTGAAGAGAAATGATTAACAATTTGAGCAACCCCACTGAT 1080
 1021 GTGCGGAGCAGGCTTGAAGAGAAATGATTAACAATTTGAGCAACCCCACTGAT 1080
 1021 GTGCGGAGCAGGCTTGAAGAGAAATGATTAACAATTTGAGCAACCCCACTGAT 1080
 1081 CTCGCCCTCTTTGAGAGAGATGAAATGCGCAGGCAAGAGGTTGCTGACACA 1140
 1081 CTCGCCCTCTTTGAGAGAGATGAAATGCGCAGGCAAGAGGTTGCTGACACA 1140
 1081 CTCGCCCTCTTTGAGAGAGATGAAATGCGCAGGCAAGAGGTTGCTGACACA 1140
 1141 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1200
 1141 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1200
 1141 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1200
 1201 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1260
 1201 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1260
 1201 CAGCTGTCAGAGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1260
 1261 ATGATTCGAGATGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGA 1320
 1261 ATGATTCGAGATGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGA 1320
 1261 ATGATTCGAGATGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGA 1320
 1321 GAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1380
 1321 GAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1380
 1321 GAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1380
 1381 CAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1440
 1381 CAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1440
 1381 CAGCTTGAATGATTTCTGATGAGATGATGATTAATGCTTGAAGAGAGA 1440
 1441 ATGCGAGGAAAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1500
 1441 ATGCGAGGAAAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1500
 1441 ATGCGAGGAAAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1500
 1501 AAAAGGATCAGATGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1560
 1501 AAAAGGATCAGATGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1560
 1501 AAAAGGATCAGATGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1560
 1561 GGAAGGTTTGGGTGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1620
 1561 GGAAGGTTTGGGTGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1620
 1561 GGAAGGTTTGGGTGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1620
 1621 TCAATGATCTAGATGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1680
 1621 TCAATGATCTAGATGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1680
 1621 TCAATGATCTAGATGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1680
 1681 TTTTAAATTTTCTTCTGATTTCTAGATTTCTAGATTTCTAGATTTCTAGAT 1740
 1681 TTTTAAATTTTCTTCTGATTTCTAGATTTCTAGATTTCTAGATTTCTAGAT 1740
 1681 TTTTAAATTTTCTTCTGATTTCTAGATTTCTAGATTTCTAGATTTCTAGAT 1740
 1741 ATGCGCAACCAAGAGCTGCTGAGATTTCTGAGATTTCTGAGATTTCTGAGAT 1800
 1741 ATGCGCAACCAAGAGCTGCTGAGATTTCTGAGATTTCTGAGATTTCTGAGAT 1800
 1741 ATGCGCAACCAAGAGCTGCTGAGATTTCTGAGATTTCTGAGATTTCTGAGAT 1800
 1801 TCTATGCTTCAAGGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1860
 1801 TCTATGCTTCAAGGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1860
 1801 TCTATGCTTCAAGGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 1860
 1861 GGCATTTTGAAGGCGCAAGCGCTTCAACCGGAGTTAGTGAATGAGAGATTT 1920
 1861 GGCATTTTGAAGGCGCAAGCGCTTCAACCGGAGTTAGTGAATGAGAGATTT 1920

QY	1921	ACCTGGGGGTCAAGAACCGCAGATATGTTTTCTGGAAGGTGGCCCAACTGATAGGCTCCTT	1990
Dp	1921	GCCTGGGGGTCAAGAACCGCAGATATGTTTTCTGGAAGGTGGCCCAACTGATAGGCTCCTT	1980
QY	1981	GCAACGAGCAGAATCTTGACATATTCATCAACACTCTCAAGGTTCTCTCCCCAAGA	2040
Dp	1981	GCAACGAGCAGAATCTTGACATATTCATCAACACTCTCAAGGTTCTCTCCCCAAGA	2040
QY	2041	AATTTGATATATGCTTTTATGTTTTGTCATGTGAATTAAAGTTTTGTTGTCGGTGTAA	2100
Dp	2041	AATTTGATATATGCTTTTATGTTTTGTCATGTGAATTAAAGTTTTGTTGTCGGTGTAA	2100
QY	2101	TGCAATCTGTTATGTAATATATCTATGATTCATTTAGGCCAAACAAGCTGAAATTTGAGTGG	2160
Dp	2101	TGCAATCTGTTATGTAATATATCTATGATTCATTTAGGCCAAACAAGCTGAAATTTGAGTGG	2160
QY	2161	AAATCATATCCAAAGAGATGGTGTGTAAGGAGCTGAGGCGAGATCTCTGAGGACAAATCAGCAG	2220
Dp	2161	AAATCATATCCAAAGAGATGGTGTGTAAGGAGCTGAGGCGAGATCTCTGAGGACAAATCAGCAG	2220
QY	2221	CTGAACATACTTTAAGAACCAAGCTCTCAAAACAGAACCAAGCAGCAAGGTCCTTGAGGAA	2280
Dp	2221	CTGAACATACTTTAAGAACCAAGCTCTCAAAACAGAACCAAGCAGCAAGGTCCTTGAGGAA	2280
QY	2281	TCTCTGGAAATTTATGAGCGAGAAAGCTGGCTGTGAATCTGACAGGATATCGATCTGTAGA	2340
Dp	2281	TCTCTGGAAATTTATGAGCGAGAAAGCTGGCTGTGAATCTGACAGGATATCGATCTGTAGA	2340
QY	2341	CAGAGAACTAAGATGACGACATGAACAGAACAGGGGAGGATATGTTTTCTCTGAAAAAT	2400
Dp	2341	CAGAGAACTAAGATGACGACATGAACAGAACAGGGGAGGATATGTTTTCTCTGAAAAAT	2400
QY	2401	CACAAACTTGAACATTTGTATTTACCTACTGATTCACATTTTGTATATATGTCACAA	2460
Dp	2401	CACAAACTTGAACATTTGTATTTACCTACTGATTCACATTTTGTATATATGTCACAA	2460
QY	2461	AAAACTGTGTGGTTTGAAGATGATGACACACGACAGTTTTCATGATTTCAATCAA	2520
Dp	2461	AAAACTGTGTGGTTTGAAGATGATGACACACGACAGTTTTCATGATTTCAATCAA	2520
QY	2521	CAGATCCATGAAGAGAGAGCGCAAGAGAGAAATTCGAGATGTTGCACACACAGGAA	2580
Dp	2521	CAGATCCATGAAGAGAGAGCGCAAGAGAGAAATTCGAGATGTTGCACACACAGGAA	2580
QY	2581	CGTGCAGAGGTGTGTTGGCCACAGCAGCAGAACTTAATCCCTCTAGCAATGACGATTTGC	2640
Dp	2581	CGTGCAGAGGTGTGTTGGCCACAGCAGCAGAACTTAATCCCTCTAGCAATGACGATTTGC	2640
QY	2641	CGAAAGAGGTATATGTAATACTTAACATAATATCCCTCTGGCGTTTTGTTTTCAAACTTA	2700
Dp	2641	CGAAAGAGGTATATGTAATACTTAACATAATATCCCTCTGGCGTTTTGTTTTCAAACTTA	2700
QY	2701	AGAGTAATCTGAATTAATTCGCCGTTTTGATTTCTTTGGCAGAGCTGAGGAAGTGTCAAGCTTC	2760
Dp	2701	AGAGTAATCTGAATTAATTCGCCGTTTTGATTTCTTTGGCAGAGCTGAGGAAGTGTCAAGCTTC	2760
QY	2761	ATCCAGTTTCAAGAGAAAGATGAGGAGGATTTGTGGAAGAGAGGAGAGTCTGATATAAA	2820
Dp	2761	ATCCAGTTTCAAGAGAAAGATGAGGAGGATTTGTGGAAGAGAGGAGAGTCTGATATAAA	2820
QY	2821	GATCAAGAGAGATGAGAGACATGAAAGAGGCAATCAAGAGATATTTGATCTG	2880
Dp	2821	GATCAAGAGAGATGAGAGACATGAAAGAGGCAATCAAGAGATATTTGATCTG	2880
QY	2881	GAGAAAGAAATTTGATGAGGCTTTTGGAACAGTCAATGTAACAGCATGGCTTCACAATGAA	2940
Dp	2881	GAGAAAGAAATTTGATGAGGCTTTTGGAACAGTCAATGTAACAGCATGGCTTCACAATGAA	2940
QY	2941	GATGATTTGAGCAAAAGTCTGTTACACAAGACAAGACTTAATTTCTTTGTTTTGCTTTTG	3000
Dp	2941	GATGATTTGAGCAAAAGTCTGTTACACAAGACAAGACTTAATTTCTTTGTTTTGCTTTTG	3000
QY	3001	GTATGTGCGAAAGTAGAGATCTGAGAGACTTCATTAAATATCTAGCAAAATCTAAGAA	3060

Accession	Sequence	Position
D6	GTATGCTGGAAAGTAGGAGATCTGAGAGACGTCCATTTTAAATACTAGAGCAAAATCTAAGGA	3060
OY	GATATATGATTAATAATCTCCCAATTTTATGAGCGGATCTAAGAGCATTAAGTCTT	3120
D6	GATATATGATTAATTAATCTCCCAATTTTATGAGCGGATCTAAGAGGACATTAAGTCTT	3120
OY	GTGACTAATAACCAAGTTTCTCTAGATTAATTTGTTTTTTTGGTAAATTTTCATATGAAAG	3180
D6	GTGACTAATAACCAAGTTTCTCTAGATTAATTTGTTTTTTTGGTAAATTTTCATATGAAAG	3180
OY	TTAGACATTAATACCAACGTCAGAGTAATCACAGATGCGAAATCAATCATGTTTTT	3240
D6	TTAGACATTAATACCAACGTCAGAGTAATCACAGATGCGAAATCAATCATGTTTTT	3240
OY	AGATTTTATATCTACAAAATATATTTGGTGCACAAAT	3275
D6	AGATTTTATATCTACAAAATATATTTGGTGCACAAAT	3275

RESULT 2	LOCUS	81365 bp	DNA	linear	PLIN 27-DEC-2000
AB025633	Arabidopsis thaliana genomic DNA, chromosome 5, F1 clone:MOM1.				
DEFINITION	AB025633 BAO00015				
ACCESSION	AB025633.2	GI:10178221			
VERSION					
KEYWORDS					
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (sites)				
	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabei,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones				
JOURNAL	DNA Res. 7 (1), 31-63 (2000)				
MEDLINE	20181125				
PubMed	10718197				
REFERENCE	2 (bases 1 to 81365)				
AUTHORS	Nakamura,Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
COMMENT	On Sep 15, 2000 this sequence version replaced gi:4589439. Address for correspondence: kao@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/and_graph.cgi?c=MOM1 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Graal-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremmlin.zool.iastate.edu/cgi-bin/rp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K15M13 and the 3' clone is M0011. Location/Qualifiers 1. 81365 /organism="Arabidopsis thaliana" /mol_type="genomic DNA"				

/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MQM1"
/clone_11b="Mitsui P1"
join(3184, .3403, 3474, .3595, 3681, .3723, 3777, .3898, 4042, .4198, 4445, .4515, 4601, .4714, 4888, .4988, 5300, .5420)
/note="gb|AAFO1580.1
gene_id:MQM1.24
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97251.1"
/db_xref="GI:8809710"
/translation="MHVPGTSDVESPAHKVDETPPKHVQFLBPIISKTVVDAQNP
YGSAPDPSSNSPLSPVPEEPPSSFEQDDPLPGIEDLOISGEVPGHBLQACG
YSLNGTSCNFEWVCHLEDGVSIVYIDESSERTSNAGAKQPNVLTADVDYLALIEV
QPLDDNRKGRKPRKRYDYLTVLADVTLLPSPADCGPFFMMFALFICVPLISGG
ACDPMQSNIEKTLHTGHASKVSLAVGVVIMERATLSTIREGYSIKCISDLTIAE
KFASATVTIPFGQPAELVIGSDGSHSLRADNSPDLISRDIVLTLRTKRM
FFKALSFYHFRCSCHKRGNTL"
join(6317, .6428, 6508, .6629, 6714, .6756, 6841, .6932, 7121, 7169, 7396, .7466, 7550, .7663, 7785, .7882, 8112, .8181)
/note="gb|AAFO1580.1
gene_id:MQM1.23
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97250.1"
/db_xref="GI:8809709"
/translation="MQPGTSVSPALNOADETPPISTKVAVNDTKNIPEYDDQDAFL
PALENTIAGEPFGHLOACGYSINGTSCNFEWCHLEDGVSIVYIDAKKPNVLT
ADVGLCLAEVQPLDDNRKGLVFPANDNRKICHPBMQSNIDKTLHTGHASKV
SLAIFVHWEAATLSIEREGYTKCNNDLTITEKFSASTAVKIPREKAEVLVIGSD
GSHCLRVNDWEPDIRLCEKEKRVLLVKVYVFEQKL"
complement(join(8642, .8705, 8794, .8863, 8978, .9387, 9384, .10347))
/note="gene_id:MQM1.22
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97249.1"
/db_xref="GI:8809708"
/translation="MSWMKSGSGMTAFDLKOROKOGLSEVEGDPPEPVSTVNASF
GVRLRLRNEPSEKSFSSVLLPSPRPALTENOCNORGCCCKRKPPTLSIPNS
HDLAFTLKEMNSWADNLIIRDVLSTEDFEMALFLKGMVSGKDEPPTIKICY
SSDRSEERTPEKTVTSVMAARSTFEDAGKVLNDSGSSFLVNASNEKFPDI
SELDSIORLOQSIPIEPEWEDDYLISHKDALKVMASGNSHRAQNAQORVDAHA
KQSDKAREMLAEKINAERAKKIIGITKNNDIWKLDLGHATEAVQALQERLQM
IEGFTYRVSYPNRKGRKNAALRSASOEPGRLDDEGMECORSTRELANSLQVITIG
IGKHSRQASLPVLAKTFFEDNRYRPDETRRGVITVTPKFRHS"
complement(12417, .13424)
/note="contains similarity to unknown protein
gb|AAF27018.1
gene_id:MQM1.21"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97248.1"
/db_xref="GI:8809707"
/translation="WATDSQPNOKLTPLKTRIALTVISTWTDNAQRPDGTINRFLK
LDFPAPNPKEVNIWSTSDPVVDOSRDLNRLTYPHVGDKIPVVFPHGGGAFIS
PNAYVNVCRPAKRLPAYIVSVNRLAPRKYPAQYDDGDFALKIENHGSILPA
NADLSRCFPAGDSAGNIAHVAVIRICEPSSSTFAYVLIGLISIOPEGGEETEAR
KQLVGAPLVSPTRIDWCMKAGLNRDHAUVGPNNAVDISGLDYPETMVVAGFDL
KDWRSYEWLKLCKGKATLLIYPMWFAFTIFPLPAGQILNRKIDPDERVAAISLS
A"
join(14922, .14981, 15072, .15167, 15563, .15666, 15755, .15891, 16096, .16178)
/note="gene_id:MQM1.20"
/codon_start=1
/evidence=not experimental
/product="50S ribosomal protein L24"
/protein_id="BAA97247.1"

/db_xref="GI:8809706"
/translation="MGWKAELIRHMKLIRGDNVMIIRGDKGETGIIKRVIRSONR
VTEGKULAKKHKKGPDPHEGGLPTEVAPLHNSVNVQVDPVGRCKGVKYLEDTK
VVARGTGTSSTIIPREELIKTRATRPPTAPKQTPMEFWEQYDAKTKGMPDL"
join(16823, .16952, 17143, .17320, 17412, .17493, 17675, .17851, 17943, .18110, 18199, .18390)
/note="gene_id:MQM1.19"
/codon_start=1
/evidence=not experimental
/product="26S proteasome, non-ATPase regulatory subunit"
/protein_id="BAA97246.1"
/db_xref="GI:8809705"
/translation="MERLQIRFGAGGLGHASPDFTLDSQVYISLALMKLKHG
RACVPMVGMGLGEFVDETVAVDVVFPMPQSGVSVAVDHFQTMMLKQTG
RPMVAVGMHSHPGFCMGLSVQVINTQSFALNCRAVAVDPIQSVKGVVILAFR
SINPQITMGQEPQRTSNLGHNRKSIALTHGLRHHYSIAINRYRKNLEKMLLN
LHKRWTDGLTLRFTHSKTNEQTOEMLSLAAKKNKAVQEBDELSPKLAIVNVR
QDAKHLIEHVSIMSSNIVQTLGMLDVVF"
complement(join(18965, .19043, 19240, .19298, 19419, .19536, 19666, .19739, 19908, .19961, 20387, .20530))
/note="contains similarity to unknown protein
emb|CAA22897.1
gene_id:MQM1.18"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97245.1"
/db_xref="GI:8809704"
/translation="MDKMQAIEKMKLVGMEVEDEQADDESSISFMEDLNRCAL
TTKORFYGRALCSAGICTLLSLMVAEPNVKRGITPTLGNLALGSTAFLIQRQV
TMLDPAKRYATLALYLSAIIIALFCALYVKNKLLTLLAILLEFTGLIWSLSYIPFR
TMSKRIPTCEDEIE"
join(22672, .23634, 23708, .23999, 24111, .24355, 24458, .24624, 24715, .24925)
/note="emb|CAB62356.1
gene_id:MQM1.17
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97244.1"
/db_xref="GI:8809703"
/translation="MSSRAGPMSKEKNVQGGYREVEQLVQGLAGTLASSODGEM
EVTIRKKNKRGNTSKTWTYVNSGNPNRAPMGQGGGAVSGNNVSGNGNGG
IQANISGRGALSKRTDNNVYAPFPVSRPPLBEGMMQAGGSAQHTAVQEPVEDD
VDNASEBNDSDALDDSDDDLASDYSDVSQKSHGRKQKWKPKKRGSLDSISTE
INBPQRWCHCAPQNGPGLIDVYNLHPLLAIRKARVLRVHLELVEFKDQMG
ASVIECGEIIYQOMKGLGDEKDYBIIVPWPVIMNTLRDKDNDKMLKMGVQELIEF
DKYELRABHSGYQGHGMSVLMFESAGTGLAEERLHRELAVELGLRLAMQGRSM
PSGVRQIYGLATRKODIDINQHSQGTGLKFLKFSJOEYVRELKQISDNQQLY
FNKLSKQNKAKVULESELTMSKLRRTAEDNRIVNORTMOHOREEMDADREF
MDSIKQIHERDPAKEENBEMLQOQERAVVQCOONINPSNDNCCRRAREVSFTIEP
QEKEMEERFEREMLIKQEKEMDMKRRHHEELFDEKEPDEALBQLMYHGLGHAND
D"
join(25614, .25838, 26061, .26207, 26324, .26438, 26516, .26610, 26710, .26813, 27522, .27765, 27943, .28036, 28115, .28305, 28385, .28604, 28688, .28855, 28944, .29092)
/note="gb|AAC97420.1
gene_id:MQM1.16
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db_xref="GI:8809702"
/translation="WAPPAGQTAVAEAAAGDAPPOOORGGSTISGIVRIAYWYF
ASKFSPKQKPMQPLAQMLNLPHKESLDMMWYLSQEKENDPQNDPALMYHETN
IPAAWTPESIRKSLIYPSPTLQNNLSLVAHIFPAGSGPFDITDPEIOPUNFSR
THAVATYFPKOKKKKSLGSPKDSSEPEVEKVGDKSPKSEVPEWLSLWKN
VTINLVDPDTPQADLIVPTTGNVYPTIYFNELRDKFIPVNETVELPLNLEISF
ISMKQQLQOQDQSFQWORSYGMWLDGESDLSKVPLEGNGYLLGITMFVMSHSVF
DLAFKNDIQFMNKNKSWGSAKSVYALNQPATIPYLILNDNSMTLASGVQVC
IEFWKIGKAMRIEVRSGMIPLRPHDRSIVASNTKTERDDLAIKFLSVILLIYIGL
SIYSILAEHRKMSWISLSTSCVYMPGFIIMCQPLFNYLKLKVAHLPMQOMYKFL
LNTIIDLPAFYIKMPLIHLRLSVFRDVLFIYLYQWVYVDPKTRVNEFGGDEET
AEKCLITEKEEDKKTN"

CDS		complement(join(29439..29576,29661..29885,29985..30268, 30341..30493,30579..30722,30956..31484))		/note="gene_id:WOM1.15"	
Query Match	100.0%;	Score 3275;	DB 8;	Length 81365;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 3275;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	GACAAACAACAAATTAAGCAAGTCATGTTGCTAGCAATAATTAATAGTGGGAAACA	60		
DB	21977	GACAAACAACAAATTAAGCAAGTCATGTTGCTAGCAATAATTAATAGTGGGAAACA	22036		
QY	61	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAACAATCAACTGAAT	120		
DB	22037	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAACAATCAACTGAAT	22096		
QY	121	GAATAATTTGGAGTCCAGAAATCGGAAAAAGAGCCGTTTAAAGCTTAATAGCTTC	180		
DB	22097	GAATAATTTGGAGTCCAGAAATCGGAAAAAGAGCCGTTTAAAGCTTAATAGCTTC	22156		
QY	181	ATTGTCTCTCTCGTCAGTTTATTTCTTCCGCGAGTCCGTCACCTACCTC	240		
DB	22157	ATTGTCTCTCTCGTCAGTTTATTTCTTCCGCGAGTCCGTCACCTACCTC	22216		
QY	241	TCTCCGCGCTTAACTTAAGTTCCTCGTCTCTCTGTAAGTTTCTGCTTAGAG	300		
DB	22217	TCTCCGCGCTTAACTTAAGTTCCTCGTCTCTCTGTAAGTTTCTGCTTAGAG	22276		
QY	301	CCTCCGATGCGCTCAGCCGATGCAATCTGTGCTCGAATTTCTTTTCTGCTGAAA	360		
DB	22277	CCTCCGATGCGCTCAGCCGATGCAATCTGTGCTCGAATTTCTTTTCTGCTGAAA	22336		
QY	361	ATTGCCCTAATGTTCTGAAATTTGGAAGTTTGTGCTAAGGTTACTTTTCCCTATA	420		
DB	22337	ATTGCCCTAATGTTCTGAAATTTGGAAGTTTGTGCTAAGGTTACTTTTCCCTATA	22396		
QY	421	TTTTATAGTCTTAGGTAACGATACCTGCTCTTACTGTTTGTTCATTTGTGTGCT	480		
DB	22397	TTTTATAGTCTTAGGTAACGATACCTGCTCTTACTGTTTGTTCATTTGTGTGCT	22456		
QY	481	TTTACCGTTTATGCTGCTGATCGGAGTATTGACTGTGAAAAATCCTTCTGTTTGTG	540		
DB	22457	TTTACCGTTTATGCTGCTGATCGGAGTATTGACTGTGAAAAATCCTTCTGTTTGTG	22516		
QY	541	TGTTTCAATATAAATGGAATGATCTACCTTTGTGCTTTGATGTTTGTGAGCCT	600		
DB	22517	TGTTTCAATATAAATGGAATGATCTACCTTTGTGCTTTGATGTTTGTGAGCCT	22576		
QY	601	ATGCGTGTGTGCTGTTAATCTTCACTGATGCTGATGTTGAGATTTGATGATG	660		
DB	22577	ATGCGTGTGTGCTGTTAATCTTCACTGATGCTGATGTTGAGATTTGATGATG	22636		
QY	661	ACTGTGGGTTTCTTGTGCTGCTAATAGTTGTAATAATGAGTTCTAGGGCTGCTCATGT	720		
DB	22637	ACTGTGGGTTTCTTGTGCTGCTAATAGTTGTAATAATGAGTTCTAGGGCTGCTCATGT	22696		
QY	721	CTAAGGAAAAAGACGTTCAAGGCTGTTATAGGCTGAGGCTGAGGTTGTTCAAGGTT	780		
DB	22697	CTAAGGAAAAAGACGTTCAAGGCTGTTATAGGCTGAGGCTGAGGTTGTTCAAGGTT	22756		
QY	781	TGGCAGGAGCAGACTGGCTTCTTCAAGATGATGAGAGAGAGTGGAGTCAATTCCA	840		
DB	22757	TGGCAGGAGCAGACTGGCTTCTTCAAGATGATGAGAGAGAGTGGAGTCAATTCCA	22816		
QY	841	AGAAAGAAAGAAACCAAGGAAACATTTGGAATACTTGGGTTTCTCAGAAATTCGA	900		
DB	22817	AGAAAGAAAGAAACCAAGGAAACATTTGGAATACTTGGGTTTCTCAGAAATTCGA	22876		
QY	901	ATCCTCTAGAGCTTGGGTTGCTCAGCAGCAAGGAGAGGTAGCAAGTATCTGGAGAG	960		
DB	22877	ATCCTCTAGAGCTTGGGTTGCTCAGCAGCAAGGAGAGGTAGCAAGTATCTGGAGAG	22936		
QY	961	GAACAATATATTCGGGAGAGGTAAAGGCAATGTGCGGGCATTTCAAGCTAAATATCTG	1020		

DB	22937	GAACAATATATTCGGGAGAGGTAAAGGCAATGTGCGGGCATTTCAAGCTAAATATCTG	22996		
QY	1021	GTCGGGAGAGAGCTTGGACAGAAAGTATGATTAACAATTTGTGGACACCCCACTGTAT	1080		
DB	22997	GTCGGGAGAGAGCTTGGACAGAAAGTATGATTAACAATTTGTGGACACCCCACTGTAT	23056		
QY	1081	CTCGCCCTCTTTTGAAGAGATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGACA	1140		
DB	23057	CTCGCCCTCTTTTGAAGAGATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGACA	23116		
QY	1141	CAGCTGTGAGAGATTTCTGACGTGAGAGATGATGTGATAATGCTTCTGAGAAAGACA	1200		
DB	23117	CAGCTGTGAGAGATTTCTGACGTGAGAGATGATGTGATAATGCTTCTGAGAAAGACA	23176		
QY	1201	ATGATTCGATGCTTGTGATGATTCGATGACGACCTTCGCAAGTGTATATATGATCGG	1260		
DB	23177	ATGATTCGATGCTTGTGATGATTCGATGACGACCTTCGCAAGTGTATATATGATCGG	23236		
QY	1261	ATGTAGTCAGAAAGAGCCATGATCAAGAAAGAGATTAAGTGTCAAAAAGTTCTTG	1320		
DB	23237	ATGTAGTCAGAAAGAGCCATGATCAAGAAAGAGATTAAGTGTCAAAAAGTTCTTG	23296		
QY	1321	GCAGCTTGATAGCTTGTGATCGAGCAGATTAATGAACACAGAGCAGTGGCATTTGTC	1380		
DB	23297	GCAGCTTGATAGCTTGTGATCGAGCAGATTAATGAACACAGAGCAGTGGCATTTGTC	23356		
QY	1381	CAGCTTGATAGAGCAGCCTGCTGATGATTTGGTATTAACCTGCACCTCTATAGCTTC	1440		
DB	23357	CAGCTTGATAGAGCAGCCTGCTGATGATTTGGTATTAACCTGCACCTCTATAGCTTC	23416		
QY	1441	ATGCGAGACAAAGAGCTAGGCGAGTTAAGCTCATAGAAATTTGGCTGAAGTTTAG	1500		
DB	23417	ATGCGAGACAAAGAGCTAGGCGAGTTAAGCTCATAGAAATTTGGCTGAAGTTTAG	23476		
QY	1501	AAAAGATCTACAGATGAGAGCGCATCTGATTCCTTGTGTGAGATTTAATGGCAGT	1560		
DB	23477	AAAAGATCTACAGATGAGAGCGCATCTGATTCCTTGTGTGAGATTTAATGGCAGT	23536		
QY	1561	GGAAGGTTTGGGTGAGATGAAAAAGATTATGAATTTGTCTGGCTCCAAATGGTATCA	1620		
DB	23537	GGAAGGTTTGGGTGAGATGAAAAAGATTATGAATTTGTCTGGCTCCAAATGGTATCA	23596		
QY	1621	TCATGAATCTAGACCTGATTAAGACGATPAAGGTGAATTTCTGCTCTTTAG	1680		
DB	23597	TCATGAATCTAGACCTGATTAAGACGATPAAGGTGAATTTCTGCTCTTTAG	23656		
QY	1681	TTCTTTAATTTTCTCTTGCAATTTCACTGATCTAGAAATGTATATGTAGTGGCTCGC	1740		
DB	23657	TTCTTTAATTTTCTCTTGCAATTTCACTGATCTAGAAATGTATATGTAGTGGCTCGC	23716		
QY	1741	ATGGGCAACAGAGCTGTGGAATATCTTCCGCAAGATATAGAGCTTTAGACAGCCAT	1800		
DB	23717	ATGGGCAACAGAGCTGTGGAATATCTTCCGCAAGATATAGAGCTTTAGACAGCCAT	23776		
QY	1801	TCCTATAGTCCACAGGCGCATGTGGAGATGATGTTTCTGATTTGAGAGCAGTGCAT	1860		
DB	23777	TCCTATAGTCCACAGGCGCATGTGGAGATGATGTTTCTGATTTGAGAGCAGTGCAT	23836		
QY	1861	GGCTATTTTGGAGCCGAAAGCCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAAT	1920		
DB	23837	GGCTATTTTGGAGCCGAAAGCCTCCACCGGAGTTAGCTGAGATGGGTTAGATGAAT	23896		
QY	1921	GGCTGGGGTCAGAACGCAATATGTTTCTGAGAGTGTTCGCAACTGATATGGCTTCTT	1980		
DB	23897	GGCTGGGGTCAGAACGCAATATGTTTCTGAGAGTGTTCGCAACTGATATGGCTTCTT	23956		
QY	1981	GCAAGAGCAAGACTGTGACATATTCATCAACTCTCAAGGTTCTCTCCCAAGAGA	2040		
DB	23957	GCAAGAGCAAGACTGTGACATATTCATCAACTCTCAAGGTTCTCTCCCAAGAGA	24016		
QY	2041	AATTTGATATATGCTTTTATGTTTGTCAATGGAATTTAAGTTTGTGTGCTCGTGTAA	2100		

```

Db      24017 AATTGATATATGCTTTAGTTTGTGTCATGGAATTTAAAGTTTGTGTCCTGTTAA 24076
Qy      2101 TGCATCTGTATGTATATATCTATGATTCATAGGCAAAACAGGCTGAAATTCGAGTTG 2160
Db      24077 TGCATCTGTATGTATATATCTATGATTCATAGGCAAAACAGGCTGAAATTCGAGTTG 24136
Qy      2161 AATCATACCAAGAGATGTTGTAAAGAGCTGAGGAGATCTTGAGAGACATCAGCAG 2220
Db      24137 AATCATACCAAGAGATGTTGTAAAGAGCTGAGGAGATCTTGAGAGACATCAGCAG 24196
Qy      2221 CTGAATCTATTATAGAACAGCTCTCAAAACAGAACAGGCAAGGCTGAGGAA 2280
Db      24197 CTGAATCTATTATAGAACAGCTCTCAAAACAGAACAGGCAAGGCTGAGGAA 24256
Qy      2281 TCTCTGAAATTTATGAGCAGAGAGCTGCTAGAACCTGACAGAGATATGAGATCGTGAGA 2340
Db      24257 TCTCTGAAATTTATGAGCAGAGAGCTGCTAGAACCTGACAGAGATATGAGATCGTGAGA 24316
Qy      2341 CAGAGAACTAAGATGACAGATGAAACAGAACAGGAAAGAGATGATTTTCTAGAAAT 2400
Db      24317 CAGAGAACTAAGATGACAGATGAAACAGAACAGGAAAGAGATGATTTTCTAGAAAT 24376
Qy      2401 CAGAACTTGAACATTTTGTATTAACCTAGATTCACATTTTGTATATATGCTCAACA 2460
Db      24377 CAGAACTTGAACATTTTGTATTAACCTAGATTCACATTTTGTATATATGCTCAACA 24436
Qy      2461 AAAACCTGTGTTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
Db      24437 AAAACCTGTGTTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 24496
Qy      2521 CAGATCCATGAAAGAGAGACGCAAGAGAGAGATTTGAGATGTTTGAAGAGAGAGAA 2580
Db      24497 CAGATCCATGAAAGAGAGACGCAAGAGAGAGATTTGAGATGTTTGAAGAGAGAGAA 24556
Qy      2581 CGGCGCAAGTTGTTGGCCAGCAGCAGCAGACATTAATCCCTCTAGCATGACGATTCG 2640
Db      24557 CGGCGCAAGTTGTTGGCCAGCAGCAGCAGACATTAATCCCTCTAGCATGACGATTCG 24616
Qy      2641 CGAAGAGATATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2700
Db      24617 CGAAGAGATATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 24676
Qy      2701 AGAGTAACTGAATATCCGCTTTGATCTCTTTCGAGAGCTGAGAAAGTCAAGCTTC 2760
Db      24677 AGAGTAACTGAATATTCGCTTTGATCTCTTTCGAGAGCTGAGAAAGTCAAGCTTC 24736
Qy      2761 ATCAAGTTTCAAGAGAGAGATGAGAGAGTTTGTGAGAGAGAGAGATGCTGATAAA 2820
Db      24737 ATCAAGTTTCAAGAGAGAGATGAGAGAGTTTGTGAGAGAGAGAGATGCTGATAAA 24796
Qy      2821 GATCAAGAGAGAGATGAGAGAGATGAGAGAGAGATGAGAGAGATGAGATGATGATG 2880
Db      24797 GATCAAGAGAGAGATGAGAGAGATGAGAGAGAGATGAGAGAGATGAGATGATGATG 24856
Qy      2881 GAGAGAGAGATTTGATGAGAGCTTTGAGACAGCTCATGTAACAAGATGAGCTTCACATGAA 2940
Db      24857 GAGAGAGAGATTTGATGAGAGCTTTGAGACAGCTCATGTAACAAGATGAGCTTCACATGAA 24916
Qy      2941 GATGATGAGACAAAGATCTGTAACAAGACAAAGATGATGATGATGATGATGATGATG 3000
Db      24917 GATGATGAGACAAAGATCTGTAACAAGACAAAGATGATGATGATGATGATGATGATG 24976
Qy      3001 GTATGTCGAGAAAGTATGAGATCTGAGAGATCTCAATTTAAATCTAGAGCAAAATCTAAGGA 3060
Db      24977 GTATGTCGAGAAAGTATGAGATCTGAGAGATCTCAATTTAAATCTAGAGCAAAATCTAAGGA 25036
Qy      3061 GATTATATGATTAATTCCTCCAAATTTTATAGAGAGATCTAAGAGAGATTAAGTTCTT 3120
Db      25037 GATTATATGATTAATTCCTCCAAATTTTATAGAGAGATCTAAGAGAGATTAAGTTCTT 25096
Qy      3121 GTGACTAAACCAAGATTTCTTATGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 3180
Db      25097 GTGACTAAACCAAGATTTCTTATGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 25156

```

```

Qy      3181 TTAGACATATTACCAACGTCAGAGTGAATTCAGAGATGGCAATCAAAATCATGTTT 3240
Db      25157 TTAGACATATTACCAACGTCAGAGTGAATTCAGAGATGGCAATCAAAATCATGTTT 25216
Qy      3241 AGAATTTTATCTACAAATATATGTTGTTAAT 3275
Db      25217 AGAATTTTATCTACAAATATATGTTGTTAAT 25251

RESULT 3
AF239719      2254 bp      DNA      linear      PLN 02-JUN-2000
LOCUS      AF239719
DEFINITION      Arabidopsis thaliana SGS3 gene, complete cds.
ACCESSION      AF239719
VERSION      AF239719.1
KEYWORDS      GI:8164029
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2254)
Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C.,
Morel, J.B., Jonette, D., Lacombe, A.M., Nikic, S., Picault, N.,
Remoue, K., Saital, M., Vo, T.A. and Vaucheret, H.
Arabidopsis SGS3 and SGS3 genes are required for
posttranscriptional gene silencing and natural virus resistance
Cell 101 (5), 533-542 (2000)
JOURNAL
MEDLINE
PUBMED
10850495
2 (bases 1 to 2254)
Beclin, C., Mourrain, P., Vaucheret, H. and Elmayan, T.
Direct Submission
Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de
Saint-Cyr, Versailles 78026, France
FEATURES
source
1..2254
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="V"
join(<1..963,1037..1328,1440..1684,1787..1953,2044..>2254)
/product="SGS3"
join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)
/codon_start=1
/product="SGS3"
/protein_id="AA173960.1"
/db_xref="GI:8164030"
/translation="MSSRAQPMSEKKEKVVGGYREVEQLVOGLAGTRVLASQDDGEM
EVISSKNNKPKGNTSGKTNVQNSNPSPVAMGQOQGRSNVSGRNNVSGNGNGRG
IOANTSGRGLSRKYDNNFVAPVPSPLEGGQNMWRGSGSQQHTVVOEPVEDD
VNASEENDSDNLDDSDDLASDDYDSVQSGSKRQNKWFKFPGSLDSLSIQ
INRPRQWRCPCQNSGALIDMYNIRLILARTRGABRVRLHRELAIVLEKDLQMG
ASVLCGEIYGGWKGLGDEQDEYELVWPMPTIMNTRLDKDNDIMKLGMOGVELLEY
DYERALRARHSYGPQGRHSGSVLMFESSATGILELRHLRELAIEKGLDRIMQGRSM
FSGYRQLYGFATQODLIDFNQHGQTRLKFELKSYQEWVVELRQISENDQNTY
FNKLSKONKAKVLESLIEMSEKRLTAEDNRIVRQRTQHQEONRENDADHDF
MSTQIHRHRRAKEENFPMLOQGRARVVGQOONINPSSDDCRKAEREVSPIEF
QEKEMBEFVEREMLIKQEKMEKMKRHHEIFDLKEFDEALQELQMYKGLHNE
D"

ORIGIN
Query Match      68.8%; Score 2254; DB 8; Length 2254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      696 ATGAGTTCTAGGGCTGTCATGCTTAAGAGAAAGACGTTCAAGGCTGTTATAGGCT 755
Db      1 ATGAGTTCTAGGGCTGTCATGCTTAAGAGAAAGACGTTCAAGGCTGTTATAGGCT 60
Qy      756 GAGGTGAACAGTTGGTTCAAGTTTGGCAGGAGCAGACGCTGCTTCACAAAGATGAT 815

```


Db 61 GAGGTGAACAGTGGTTCAAGGTTGGCAGGAGCAGACTGGCTTCTTCAAGAATGAT 120
Qy 816 GAGAGAGAGTGGAGGTCATTTTCCAAAGAAACAACAACAACCAAGAAACATTTCTGGA 875
Db 121 GAGAGAGAGTGGAGGTCATTTTCCAAAGAAACAACAACAACCAAGAAACATTTCTGGA 180
Qy 876 AAAACTTGGGTTCTCAGAAATTCGAAATCTCTAGAGCTTGGGTGGTCAGCAGAAAGG 935
Db 181 AAAACTTGGGTTCTCAGAAATTCGAAATCTCTAGAGCTTGGGTGGTCAGCAGAAAGG 240
Qy 936 AGAGTAGCAACGTAATCTGGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 995
Db 241 AGAGTAGCAACGTAATCTGGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 300
Qy 996 CGGGGATTCAGACTAACAATCTGGTCGGGGAGAGAGCTTGGAGCAGAAAGTATGATTAAC 1055
Db 301 CGGGGATTCAGACTAACAATCTGGTCGGGGAGAGAGCTTGGAGCAGAAAGTATGATTAAC 360
Qy 1056 AACTTGTGGCAACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGGAATTTGGCAG 1115
Db 361 AACTTGTGGCAACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGGAATTTGGCAG 420
Qy 1116 GCAAGAGAGAGTCTGCTCAGACAACAAGCTGTCAGAGATTTCTGAAGTATGAT 1175
Db 421 GCAAGAGAGAGTCTGCTCAGACAACAAGCTGTCAGAGATTTCTGAAGTATGAT 480
Qy 1176 GTGGATTAATCTTCTGAGAGAGAAATGATCCGATGCTTGGATGATTTGATGACGAC 1235
Db 481 GTGGATTAATCTTCTGAGAGAGAAATGATCCGATGCTTGGATGATTTGATGACGAC 540
Qy 1236 CTTCGAAGTATGATTTATGATCGAGATGTGAGTCAAAAGAGCCATGGATCAGAAAGCAG 1295
Db 541 CTTCGAAGTATGATTTATGATCGAGATGTGAGTCAAAAGAGCCATGGATCAGAAAGCAG 600
Qy 1296 AATTAAGTGTTCAAAAGATCTTTTGGCAGCTTGGATGATCTTGTGATTCAGCAGATTAAT 1355
Db 601 AATTAAGTGTTCAAAAGATCTTTTGGCAGCTTGGATGATCTTGTGATTCAGCAGATTAAT 660
Qy 1356 GAACCAACAAGAGAGTGGCATTGTCCAGCTTGTCAAGACGACCTGGTGCATGCAATGG 1415
Db 661 GAACCAACAAGAGAGTGGCATTGTCCAGCTTGTCAAGACGACCTGGTGCATGCAATGG 720
Qy 1416 TATTAACCTGACCTCTAAGCTCATGAGGAGCAAAAGAGCTAGGGAGTAAAGCTC 1475
Db 721 TATTAACCTGACCTCTAAGCTCATGAGGAGCAAAAGAGCTAGGGAGTAAAGCTC 780
Qy 1476 CATAGAGAAATGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCATT 1535
Db 781 CATAGAGAAATGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCATT 840
Qy 1536 CCTTGTGTGAGATTTATGAGGAGTGGAGAGGTTTGGGTGAGATGAAAAAGATTTGAA 1595
Db 841 CCTTGTGTGAGATTTATGAGGAGTGGAGAGGTTTGGGTGAGATGAAAAAGATTTGAA 900
Qy 1596 AATGTCGGCTCAATGGTCATCATCATGATTAAGTAAAGTAAAGAGCAATGCAAT 1655
Db 901 AATGTCGGCTCAATGGTCATCATCATGATTAAGTAAAGTAAAGAGCAATGCAAT 960
Qy 1656 AAGGTGAATTTCTTGTCTTTTACTTCTTAAATTTTCTCTGCAATTTACTGATCTTA 1715
Db 961 AAGGTGAATTTCTTGTCTTTTACTTCTTAAATTTTCTCTGCAATTTACTGATCTTA 1020
Qy 1716 GAATGTTAATTTTATGAGTGGCTGGCAGTGGCAACAAGAGCTCTGGAATCTTCCAAA 1775
Db 1021 GAATGTTAATTTTATGAGTGGCTGGCAGTGGCAACAAGAGCTCTGGAATCTTCCAAA 1080
Qy 1776 GATAGAGGCTCTTAGAGCAGCATTCTATGTTCCACAGGGCAGATGAGGATGAGTGT 1835
Db 1081 GATAGAGGCTCTTAGAGCAGCATTCTATGTTCCACAGGGCAGATGAGGATGAGTGT 1140
Qy 1836 TCTGATGTTTGAAGAGTGCATCTGCTATTTTGAAGCCGAAACGCTCCACCGGAGTT 1895
|||||

Db 1141 TCTGATGTTTGAAGAGTGCATCTGCTATTTTGAAGCCGAAACGCTCCACCGGAGTT 1200
Qy 1896 AGCTGATGGGGTTATGATGAAATTTGCTGGGGTCAGAACCGCAGATGTTTTCTGAGG 1955
Db 1201 AGCTGATGAGGTTATGATGAAATTTGCTGGGGTCAGAACCGCAGATGTTTTCTGAGG 1260
Qy 1956 TGTTCGCAACTGATAGCTTCTTTCGCAAGAGCAAGATCTGACATATTCATCAACA 2015
Db 1261 TGTTCGCAACTGATAGCTTCTTTCGCAAGAGCAAGATCTGACATATTCATCAACA 1320
Qy 2016 CTCTCAAGTTCTCTCCCAAGAAATTTGATATATGCTTTTACTTTTGTCTGGAAT 2075
Db 1321 CTCTCAAGTTCTCTCCCAAGAAATTTGATATATGCTTTTACTTTTGTCTGGAAT 1380
Qy 2076 TTTAAGTTTGTGGCGGTATATGATCTGTTATGATATATGATATGATTTCAATTAAG 2135
Db 1381 TTTAAGTTTGTGGCGGTATATGATCTGTTATGATATATGATATGATTTCAATTAAG 1440
Qy 2136 CAAACAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAG 2195
Db 1441 CAAACAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAG 1500
Qy 2196 GCAGATCTTGAAGCAATCAGCAGCTGAACCTTTAAGAACAGCTCTCAAAACAGAA 2255
Db 1501 GCAGATCTTGAAGCAATCAGCAGCTGAACCTTTAAGAACAGCTCTCAAAACAGAA 1560
Qy 2256 CAAGCAGCAGAGTCTGAGGAATCTCGAAATTTATGAGGAGAGAGCTGCTGAAC 2315
Db 1561 CAAGCAGCAGAGTCTGAGGAATCTCGAAATTTATGAGGAGAGAGCTGCTGAAC 1620
Qy 2316 TGCAGAGATTAATCGGATGTCAGACAGAGAACTAAGATGACATGAACAGAACAGGA 2375
Db 1621 TGCAGAGATTAATCGGATGTCAGACAGAGAACTAAGATGACATGAACAGAACAGGA 1680
Qy 2376 AAGATATATTTTCTTGAAGAAATCAAACTTGCATTTTGTATTTACTTACTGATCA 2435
Db 1681 AAGATATATTTTCTTGAAGAAATCAAACTTGCATTTTGTATTTACTTACTGATCA 1740
Qy 2436 CATTTTGTATTTATTTGTCCAAACAAACACGTGTGTGTGTTGAAGATGACACAGA 2495
Db 1741 CATTTTGTATTTATTTGTCCAAACAAACACGTGTGTGTGTTGAAGATGACACAGA 1800
Qy 2496 CAGTTTTCATGATTTCAATCAACAGATCCATGAAGAAAGAGAGCAGAAAGAGAGAA 2555
Db 1801 CAGTTTTCATGATTTCAATCAACAGATCCATGAAGAAAGAGAGCAGAAAGAGAGAA 1860
Qy 2556 TTTGAGATGTTGCAGCAGCAGAAAGTGCAGAGTTGTTGGCCAGCAGCAGCAAGCAT 2615
Db 1861 TTTGAGATGTTGCAGCAGCAGAAAGTGCAGAGTTGTTGGCCAGCAGCAGCAAGCAT 1920
Qy 2616 TAACTCCCTTAGCAATGACATTTGCCGAAAGAGTTATGTAATTAATTAATCCCT 2675
Db 1921 TAACTCCCTTAGCAATGACATTTGCCGAAAGAGTTATGTAATTAATTAATCCCT 1980
Qy 2676 CTGGCGTTTGTGTTTTCAAACCTAAGAGTAACTGAATTAATCCGGTTTGTGATCTTTCG 2735
Db 1981 CTGGCGTTTGTGTTTTCAAACCTAAGAGTAACTGAATTAATCCGGTTTGTGATCTTTCG 2040
Qy 2736 CAGAGCTGAGAAAGTGTCAAGCTTCATCGAGTTTCAAGAGAAAGAGATGAGAGGTTGT 2795
Db 2041 CAGAGCTGAGAAAGTGTCAAGCTTCATCGAGTTTCAAGAGAAAGAGATGAGAGGTTGT 2100
Qy 2796 GGAAGAGAGAGAGTCTGATTAAGATCAAGAGAAAGAGATGAGAGATGAGAGAGAG 2855
Db 2101 GGAAGAGAGAGAGTCTGATTAAGATCAAGAGAAAGAGATGAGAGATGAGAGAGAG 2160
Qy 2856 GCATCAAGAGATATTTGATCTGGAAGAAATTTTGAAGGCTTTGGAACAGCTCAT 2915
Db 2161 GCATCAAGAGATATTTGATCTGGAAGAAATTTTGAAGGCTTTGGAACAGCTCAT 2220
Qy 2916 GTACAAGCATGGCTTCAATGAAGATGATGA 2949
Db 2221 GTACAAGCATGGCTTCAATGAAGATGATGA 2254
|||||

RESULT 4
BT002944
LOCUS
DEFINITION
Arabidopsis thaliana clone RAF14-93-K05 (R20243) unknown protein
(AT5G23570) mRNA, complete cds.
ACCESSION
BT002944
VERSION
BT002944.1
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosida II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 2162)
REFERENCE
AUTHORS
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
TITLE
Arabidopsis Full Length cDNA Clones
JOURNAL
Unpublished
REFERENCE
AUTHORS
2 (bases 1 to 2162)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
TITLE
Direct Submission
JOURNAL
Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
COMMENT
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinzaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Chan, M.M.,
Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S.,
Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G.,
Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M.,
Palm, C.J., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.
Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.
FEATURES
Source
1. .2162
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/note="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1. .2162
/gene="At5g23570"
1. .91
/gene="At5g23570"
92. .1969
/gene="At5g23570"
/codon_start=1
gene
S'UTR
CDS

ORIGIN
Query Match 29.7%; Score 972; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
misc_difference 2146
/gene="At5g23570"
/note="not present in genomic sequence"
3'UTR
misc_difference 2146
/gene="At5g23570"
/note="not present in genomic sequence"
/evidence=experimental
/product="unknown protein"
/protein_id="AA022757.1"
/db_xref="GI:27754623"
/translation="MSRPAQPMSEKRVVQGYRPEVQVYQGLATPTLASSODNGEM
EVISRKNKRNKRGNTSGTWTWSQNSNPRAAGQOQGGSVNSVGRNVSGRKNNGG
IQANISGRGRALSRKYDNNFVAPEPVSRPLEGGMNQAGGSAQAVAEFPEVED
VDNASFEENSDALDDSDDDLASDDYDSQKSGSRKQKMFKEFFGSDLSISIQ
INEROROMHCPACONGGALDWMYHLPLHLARTKGRARVYKLEHLEALEKDLQMG
ASVTPCEIYVQWKGGLGEDEKDYETIWPMPVITIMNTRLDNDKMTGLAMQGRSM
DKYALAPRHSYCPQGRHGSVLMFEBSATGYLEAERLHELEMEGLDRLAMQGRSM
FSGVRLDYGLATKQDLDFNHSQKRLKLELKYQEMVKEKLQISEDNQILY
PKNTLSQNKRAKXVLESLEIMSRLRTLEDNETVORTMOHEORRENDADRF
MDSIKQIHREDAENFEMLOQQRKRVVQOQONIPNSDCCRRAEVSFIEF
QEKMEHFVERREWLIKQEKMEKMKRHIEIFILEKEFDEALQIMYHGHLEHD
D"
1970. .2162
/gene="At5g23570"
/note="not present in genomic sequence"
Query Match 29.7%; Score 972; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
misc_difference 2146
/gene="At5g23570"
/note="not present in genomic sequence"
1347 CAGATTAATGAACACAGAGGCGATGGCACTTGTCCAGCTTGCAGAACGCACTGGTGC 1406

Db 743 CAGATAAATGACCAAGAGGAGGAGTTCAGCTTTCAGAAACGACCTGTGCC 802
 QY 1407 ATGATTTGGATATACCTGCACCCCTTACTAGCTCATGCGAGCAAAAGGAGCTAGGCCA 1466
 Db 803 ATGATTTGGATATACCTGCACCCCTTACTAGCTCATGCGAGCAAAAGGAGCTAGGCCA 862
 QY 1467 GTTAACCTCCATAGAGAAATTTGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCA 1526
 Db 863 GTTAACCTCCATAGAGAAATTTGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCA 922
 QY 1527 TCTGTATCTCTTGTGCTGAGATTTATGCGCAGTGAAGGTTTGGGTGAGATGAAG 1586
 Db 923 TCTGTATCTCTTGTGCTGAGATTTATGCGCAGTGAAGGTTTGGGTGAGATGAAG 982
 QY 1587 GATTATGAATTTCTGCGCTCCAAATGCTCATCATCTGAATCTAGCTGATTAAGAC 1646
 Db 983 GATTATGAATTTCTGCGCTCCAAATGCTCATCATCTGAATCTAGCTGATTAAGAC 1042
 QY 1647 GATAACGATTAAG 1658
 Db 1043 GATAACGATTAAG 1054

RESULT 5
 AX078761 1878 bp DNA linear PAT 22-FEB-2001
 LOCUS Sequence 2 from Patent WO0105951.
 DEFINITION AX078761
 ACCESSION AX078761 GI:13158380
 VERSION
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 Beclin, C., Elmayan, T. and Vaucheret, H.
 Novel sggs plant gene and use thereof
 Patent: WO 0105951-A 2 25-JAN-2001
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)

FEATURES

source

Location/Qualifiers
 1..1878
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 1..1878
 /note="unassigned protein product"
 /codon_start=1
 /protein_id="CAC32419.1"
 /db_xref="GI:13158381"
 /translation="MSSRAGPMSKKNVQGYRPEVQLVQGLRLASDDGSEW
 EVSKKNKPKGNTSGKTWSONSPRPMGQOQSGVNSGRGNVSGRGNRG
 IOANISGRGALSRLKYDNNFVAPVSRPLEEGMAMQARCAOHTAVQEPVDD
 VONASRENDSDALDDSDDDLDSDVDSDVSKSHGRKPKKPFPSLSDSTEO
 INEPQRMHCRAQONCPGALIDYNIHPLAHARTKARRYKHLRELAIVLEKQMRG
 ASVTGCEITGQWKGJSEDEKDYIWPVPIITWTRLDKDDDKMLGNGNOLLEIF
 DKTEALRAHSYVPOGHRMSVLMFESSATGLAERLHRELAEMLDRJAMQKSM
 FSGGVQLVFLATKODLDIFNQHSQKRLKFLKLSYOMVVEKLEIOLSEDOQAN
 FKRLSKNGHAKVLEESLEINSEKLRTEENRIIVORTKQHEONREMDHDEFF
 MOSIKOIHERRDAKENFEMLOQOERAKVVGQOQONINSNDGCRRAEVSSFTFF
 QEKEMEFVEREMLIKDEKEMDKKRNHEIFPLEKEFPDLALQMLTKHGLNED
 D"

ORIGIN

Query Match 29.4%; Score 963; DB 6; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 696 ATAGATTCTAGGCTGTGCTCAATGTTCTAGAGAAAGAAAGCTTACGGGTGTTATAGGCTT 755
 Db 1 ATAGATTCTAGGCTGTGCTCAATGTTCTAGAGAAAGAAAGCTTACGGGTGTTATAGGCTT 60

QY 756 GAGTTGAACAGTTGGTTCAAGTTTGGCAGGAGACAGACTGGCTTCTTCAAGATGAT 815
 Db 61 GAGTTGAACAGTTGGTTCAAGTTTGGCAGGAGACAGACTGGCTTCTTCAAGATGAT 120
 QY 816 GAGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAACAAACAGAGAAACCTTCTGA 875
 Db 121 GAGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAACAAACAGAGAAACCTTCTGA 180
 QY 876 AAACTTGGGTTTCTCAGAAATTCGAATCCTCTAGAGCTTTGGGGTGTGACAGCAAGG 935
 Db 181 AAACTTGGGTTTCTCAGAAATTCGAATCCTCTAGAGCTTTGGGGTGTGACAGCAAGG 240
 QY 936 AGAGTAGCAAGTATCTGGAGAGAGAAACAATGTATCCGAGAGAGTAAACGCAATGT 995
 Db 241 AGAGTAGCAAGTATCTGGAGAGAGAAACAATGTATCCGAGAGAGTAAACGCAATGT 300
 QY 996 CGGGGCAATCAAGCTAATCTGTGTCGGGAGCAGCGTTGAGCAGAAAGTATGATAAC 1055
 Db 301 CGGGGCAATCAAGCTAATCTGTGTCGGGAGCAGCGTTGAGCAGAAAGTATGATAAC 360
 QY 1056 AACTTTGTGCAACCCCACTGTATCTCGCCTCTTGGAAAGAGATGGAATTTGGCAG 1115
 Db 361 AACTTTGTGCAACCCCACTGTATCTCGCCTCTTGGAAAGAGATGGAATTTGGCAG 420
 QY 1116 GCAAGAGAGGTTCTGCTCAGACACACGCTGACAGAGATTTCTGACGTTGAGATGAT 1175
 Db 421 GCAAGAGAGGTTCTGCTCAGACACACGCTGACAGAGATTTCTGACGTTGAGATGAT 480
 QY 1176 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGAC 1235
 Db 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGAC 540
 QY 1236 CTTCGAGAGATGATTTATGATCTCGATGTGATGATCAAAAGCCATGATCAAGAAAGC 1295
 Db 541 CTTCGAGAGATGATTTATGATCTCGATGTGATGATCAAAAGCCATGATCAAGAAAGC 600
 QY 1296 AATTAAGTGTCAAAAAGTTCTTTGGCAGCTTGATGATGCTTGTGATGACAGATTAAT 1355
 Db 601 AATTAAGTGTCAAAAAGTTCTTTGGCAGCTTGATGATGCTTGTGATGACAGATTAAT 660
 QY 1356 GAACCAAGAGGAGTGTGATTCAGCTTGTGACAAACGACCTGTGATGATGATGATG 1415
 Db 661 GAACCAAGAGGAGTGTGATTCAGCTTGTGACAAACGACCTGTGATGATGATGATG 720
 QY 1416 TATAACCTGACCTTCTACTAGCTCATGCGAGAGCAAAAGAGCTTAGGAGTTAAGCTC 1475
 Db 721 TATAACCTGACCTTCTACTAGCTCATGCGAGAGCAAAAGAGCTTAGGAGTTAAGCTC 780
 QY 1476 CATAGAGATTTGGCTGAATTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGCATT 1535
 Db 781 CATAGAGATTTGGCTGAATTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGCATT 840
 QY 1536 CTTTGTGTGAGATTTATGAGGAGTGAAGGGTTTGGGTGAGAGTGAAGAAATTAAGA 1595
 Db 841 CTTTGTGTGAGATTTATGAGGAGTGAAGGGTTTGGGTGAGAGTGAAGAAATTAAGA 900
 QY 1596 ATTGTTCGCTCCATATGCTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 1655
 Db 901 ATTGTTCGCTCCATATGCTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 960
 QY 1656 AAG 1658
 Db 961 AAG 963

RESULT 6
 BT004380 1909 bp mRNA linear PLN 14-FEB-2003
 LOCUS Arabidopsis thaliana clone U20243 unknown protein (At5g2570) mRNA,
 DEFINITION complete cds.
 ACCESSION BT004380
 VERSION BT004380.1 GI:28393932

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT
JOURNAL
TITLE
JOURNAL
COMMENT

FLI CDNA.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1909)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 1909)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : "RIKEN
Arabidopsis Full-length cDNA": Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinzaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,
Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP
/PGEC) contributed equally to this work as PIs.
Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.
Location/Qualifiers
1..1909
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U20243"
/ecotype="Columbia"
/note="This clone is in PUNI 51."
1..1909
/gene="At5g23570"
1..1878
/gene="At5g23570"
/codon_start=1
/evidence="experimental
/product="unknown protein"
/protein_id="AA042374.1"
/db_xref="GI:2839393"
/translation="MSSRAGPMSEKKNVGGYRPEVEOLVGLAGTRLASDODGEM
EVLISKNKKPKGNTSGKTNNFVAPPPVSRPREGKMMWQAGSAOHTVAOEPDVED
IQANTSGRGALSRKYDNFVAPPPVSRPREGKMMWQAGSAOHTVAOEPDVED
VDAEENDSDALDDSDDLASDDYDVSQSGSRKQKWFKPFESLDSLEED
INPBOHMCAPACONGPAIDWYNIHLPLAHARTGARVRKQNLKRELAEVLEKDLQMRG
ASTVPGETVYGMKLGDEKDEKIVPMVLIIMTRLDKDNKMLAGVNDLEIEYF
DKYBALRAHRSYSPQCHRGMSVYLMFESSATYLAERLRELAEKGLDILANGQKRM
PSSGVOLYGFATKODLDLFNDSOGKTRKLEKSYQEMVYVVKELRQISDNQQLNY

ORIGIN
3' UTR
Query Match 29.4%; Score 963; DB 8; Length 1909;
Best Local Similarity 100.0%; Pred.No. 0; Mismatches 0; Gaps 0;
Matches 963; Conservative 0; Indels 0; Gaps 0;
D
1879..1909
/gene="At5g23570"
FNKLKONKHAQVLEESLEIMSEKLRRTAEDNRIVRQRTKQHEQNEENDADHREFF
MSXIQIHRPRDPAKEENEMILQOERARVGOOONINPSSNDCKRAEVSFIEF
QEKEMEEFVEEREMILKQEKEMEDMKRHHIEIFDLKEPDEALEQIMYKGLHNEED
D
1879..1909
/gene="At5g23570"
696 ATGAGTCTAGAGGCTGGTCCATATGCTTAAGAAAGAACCTTCAAGGCTGTTATAGGCT 755
1 ATGAGTCTAGAGGCTGGTCCATATGCTTAAGAAAGAACCTTCAAGGCTGTTATAGGCT 60
756 GAGGTTGAACAGTTGGTTCAAGGTTTGGCGAGGACGAGCTGCTTCTCAAGATGAT 815
61 GAGGTTGAACAGTTGGTTCAAGGTTTGGCGAGGACGAGCTGCTTCTCAAGATGAT 120
816 GGAGGAGAGTGGAGGCTCATTTCCAGAGAAACAGAACCAACAGAAACACTTTCGA 875
121 GGAGGAGAGTGGAGGCTCATTTCCAGAGAAACAGAACCAACAGAAACACTTTCGA 180
876 AAAAATTGGGTTTCTCAAGATTCCTTCAAGCTTGGGCTGTCAGCAAGATGAT 935
181 AAAAATTGGGTTTCTCAAGATTCCTTCAAGCTTGGGCTGTCAGCAAGATGAT 240
936 AGAGTGAACAGTATCTGGGAGGAGAAACATATATCCGGAGAGGTTAACCGCATGAT 995
241 AGAGTGAACAGTATCTGGGAGGAGAAACATATATCCGGAGAGGTTAACCGCATGAT 300
996 CGGGGCAATTCAGCTTAACATATCTGGGAGGAGGAGGCTTGAACAAAGTATGATAC 1055
301 CGGGGCAATTCAGCTTAACATATCTGGGAGGAGGAGGCTTGAACAAAGTATGATAC 360
1056 AACTTTTGGGACACCCCACTGATCTGCGCTTGGGAGGAGGATGATTTGGCAG 1115
361 AACTTTTGGGACACCCCACTGATCTGCGCTTGGGAGGAGGATGATTTGGCAG 420
1116 GCAGAGAGAGGTTCTGCTCAGACACAGCTGTGACAGAGTTCTGACGTGGAGATGAT 1175
421 GCAGAGAGAGGTTCTGCTCAGACACAGCTGTGACAGAGTTCTGACGTGGAGATGAT 480
1176 GTGGATATGCTTCTGAGAGAGAGATGATTCGATGATTTGATGATGATGATGATGAT 1235
481 GTGGATATGCTTCTGAGAGAGAGATGATTCGATGATTTGATGATGATGATGATGAT 540
1236 CTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
541 CTTGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
1296 AATAGTGGTTCAAAAAGTTCTTGGCAGCTTGAATGATGATGATGATGATGATGATGAT 1355
601 AATAGTGGTTCAAAAAGTTCTTGGCAGCTTGAATGATGATGATGATGATGATGATGATGAT 660
1356 GAACCAAGAGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
661 GAACCAAGAGAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
1416 TATTAACCTGACCTCTACTAGCTCATGCGAGAGCAAAAAGAGCTTGGGAGTTAAGCTC 1475
721 TATTAACCTGACCTCTACTAGCTCATGCGAGAGCAAAAAGAGCTTGGGAGTTAAGCTC 780
1476 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGAT 1535
781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGCATCTGAT 840
1536 CTTGTGTGATGATTTATGAGGAGTGAAGAGGTTTGGTGAAGATGATGATGATGATGATGAT 1595
841 CTTGTGTGATGATTTATGAGGAGTGAAGAGGTTTGGTGAAGATGATGATGATGATGATGAT 900
1596 ATTGTGGGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655

|||||
Db 901 ATTGTCGCGCTCATGATCATCATCAATCACTAGACTGATAGACGATAACGAT 960
QY 1656 AAG 1658
|||
Db 961 AAG 963
|||
RESULT 7
ATH528171
LOCUS 650 bp DNA linear PLN 29-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone
157C08.
ACCESSION AJ528171 GI:26796431
VERSION left border; T-DNA flanking sequence.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS 1 Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chauvin, S., Bechold, N., Cruaud, C., Derose, R., Pelletier, G.,
Lepintec, L., Caboche, M. and Lecharny, A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE 22363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 650)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.infobiogen.fr).
Location/Qualifiers
FEATURES
source
1..650
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/culivar="Wassiljewskij4"
/db_xref="taxon:3702"
/clone="157C08"
misc_feature 1..650
/note="T-DNA flanking sequence
left border"
ORIGIN
Query Match 16.5%; Score 541; DB 8; Length 650;
Best local Similarity 99.8%; Pred. No. 1.9e-296;
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1428 CCTCTACTGCTCATGCGAGACAAAAGAGCTAGCGGAGTTAGCTCCATAGAGATTG 1487
|||
Db 1 CCTCTACTGCTCATGCGAGACAAAAGAGCTAGCGGAGTTAGCTCCATAGAGATTG 60
|||
QY 1488 GCTGAAGTTTGAAGAAGATCTACAGATGAGAGGGGAGTCTGCTCTGTTGAG 1547
|||
Db 61 GCTGAAGTTTGAAGAAGATCTACAGATGAGAGGGGAGTCTGCTCTGTTGAG 120
|||
QY 1548 ATTATGGCGAGTGAAGGGTTGGTGAAGATGAAGAAGATTATGCTCTGCGCT 1607
|||

|||||
Db 121 ATTATGGCGAGTGAAGGGTTGGTGAAGATGAAGAAGATTATGCTCTGCGCT 180
QY 1608 CCAATGTCATCATGATGAATACTAGACTGATAAGACGATTAAGATAGGTGAAATTC 1667
|||
Db 181 CCAATGTCATCATGATGAATACTAGACTGATAAGACGATTAAGATAGGTGAAATTC 240
|||
QY 1668 TTCTGCTTTTAACTCTTTAAATTTTCTTTGCAATTTACTGATCTTAAAGATTACAT 1727
|||
Db 241 TTCTGCTTTTAACTCTTTAAATTTTCTTTGCAATTTACTGATCTTAAAGATTACAT 300
|||
QY 1728 GTATGCTGCTGGCAGTGGCAACCAAGAGCTGCTGGAATCTTGCACAGATAGAGCTCT 1787
|||
Db 301 GTATGCTGCTGGCAGTGGCAACCAAGAGCTGCTGGAATCTTGCACAGATAGAGCTCT 360
|||
QY 1788 TAGAGAGCGCACTTCCTATGCTCAAGGGCCATGCTGGAGATGATGTTCTGATGTTGA 1847
|||
Db 361 TAGAGAGCGCACTTCCTATGCTCAAGGGCCATGCTGGAGATGATGTTCTGATGTTGA 420
|||
QY 1848 GAGCAGTGCCTGCTATTTTGAAGCGGACACGCTCCACCGGAGTTAGCTGAGATGGG 1907
|||
Db 421 GAGCAGTGCCTGCTATTTTGAAGCGGACACGCTCCACCGGAGTTAGCTGAGATGGG 480
|||
QY 1908 GTTAAGATGAATTTGCTGGGCTCAGAGCGCATATGTTTCTGAGAGTTCGCCACT 1967
|||
Db 481 GTTAAGATGAATTTGCTGGGCTCAGAGCGCATATGTTTCTGAGAGTTCGCCACT 540
|||
QY 1968 GTATGCTTCTTGGACAGCAAGATCTGACATTTCAATCAACACTCT 2019
|||
Db 541 GTATGCTTCTTGGACAGCAAGATCTGACATTTCAATCAACACTCT 592
|||
RESULT 8
ATH527741
LOCUS 45 bp DNA linear PLN 29-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone
146F03.
ACCESSION AJ527741 GI:26796001
VERSION left border; T-DNA flanking sequence.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS 1 Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chauvin, S., Bechold, N., Cruaud, C., Derose, R., Pelletier, G.,
Lepintec, L., Caboche, M. and Lecharny, A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE 22363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 45)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.infobiogen.fr).
Location/Qualifiers
FEATURES
source
1..45
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

```

misc_feature
1. 45
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/clone="146F03"
/db_xref="taxon:3702"
/cultivar="Massillewskija"

ORIGIN
Query Match 1.4% Score 45; DB 8; Length 45;
Best Local Similarity 100.0%; Pred. No. 5-2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2033 CCCAAGAAATTTGATATATAGCTTTTGGTTTGTCTTGGAAATTT 2077
Db 1 CCCAAGAAATTTGATATATAGCTTTTGGTTTGTCTTGGAAATTT 45

RESULT 9
AC102284/c
LOCUS AC102284 170765 bp DNA linear HTG 31-OCT-2002
DEFINITION Mus musculus clone RP24-329J16, WORKING DRAFT SEQUENCE, 5 unordered
pieces
AC102284.3 GI:24431710
AC102284.3 HTG: PHASE1; HTG: DRIFT; HTG: FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 170765)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-329J16
Unpublished
2 (bases 1 to 170765)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Bana, N., Baetsen, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Gind, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagob, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacroque, K.,
Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Menelaus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Olivier, J., Peterson, K., Phunhkhong, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, N., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

```

TITLE

Peterson, K., Phunhkhong, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, N., Schuback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

COMMENT

Direct Submission
Submitted (31-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 31, 2002 this sequence version replaced gi:22381142.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submission@genome.wi.mit.edu

Project Information
Center project name: 329.J.16

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 169312 bases at least Q40
Consensus quality: 170102 bases at least Q30
Consensus quality: 170295 bases at least Q20

Insert size: 16300; agarose-fp
Insert size: 170365; sum-of-contigs
Quality coverage: 17.6 in Q20 bases; agarose-fp
Quality coverage: 16.8 in Q20 bases; sum-of-contigs

NOTES: This is a 'working draft' sequence. It currently
consists of 5 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 76853: contig of 76853 bp in length
* 76854 76953: gap of 100 bp
* 76954 78949: contig of 1996 bp in length
* 78950 92772: contig of 13723 bp in length
* 92773 92872: gap of 100 bp
* 92873 150873: contig of 58001 bp in length
* 150874 150974: gap of 100 bp
* 150974 170765: contig of 19792 bp in length.

FEATURES

source

1. 170765
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RP24-329J16"
/clone="RP24-329J16" Male Mouse BAC"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

1. 76853
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
76954..78949
/note="assembly_fragment"
79050..92772
/note="assembly_fragment"
92873..150873
/note="assembly_fragment"
150974..170765
/note="assembly_fragment"
clone_end:T7
vector_side:right"

Query Match 0.7%; Score 24; DB 2; Length 170765;
 Best Local Similarity 100.0%; Pred. No. 0.51;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAAACAAACAAATTAAGCAAG 25
 |||||
 Db 65301 ACAAAACAAACAAATTAAGCAAG 65278
 |||||

RESULT 10
 AF057056/c 554 bp DNA linear INV 19-DEC-2003
 LOCUS AF057056/c
 DEFINITION Uroleucon erigeronensis NADH dehydrogenase subunit 1-like protein (ND1) gene, partial sequence; mitochondrial gene for mitochondrial product.
 ACCESSION AF057056
 VERSION AF057056.1 GI:8745205
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Uroleucon.
 REFERENCE
 AUTHORS Moran,N.A., Kaplan,M.E., Gelsey,M.J., Murphy,T.G. and Scholes,E.A.
 TITLE Phylogenetics and evolution of the aphid genus Uroleucon based on mitochondrial and nuclear DNA sequences
 JOURNAL Syst. Entomol. 24 (1), 85-93 (1999)
 REFERENCE 2 (bases 1 to 554)
 AUTHORS Moran,N.A., Kaplan,M.E., Gelsey,M.J., Murphy,T.G. and Scholes,E.A.
 TITLE Direct Submission
 JOURNAL Submitted (03-APR-1998) Ecology and Evolutionary Biology, University of Arizona, Biological Sciences West, Tucson, Arizona 85721, USA
 FEATURES
 source 1..554
 /organism="Uroleucon erigeronensis"
 /organelle="mitochondrion"
 /mol_type="genomic DNA"
 /db_xref="taxon:87314"
 <1..>554
 /gene="ND1"
 1..554
 /gene="ND1"
 /note="similar to NADH dehydrogenase subunit 1"

ORIGIN
 Query Match 0.7%; Score 23; DB 3; Length 554;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3151 GTTTTGTGTAATTTTCAT 3173
 |||||
 Db 468 GTTTTGTGTAATTTTCAT 446
 |||||

RESULT 11
 BX004860 101080 bp DNA linear HTG 14-DEC-2002
 LOCUS BX004860/c
 DEFINITION Dantio rerio clone BUSM1-144B24, WORKING DRAFT SEQUENCE, 4 unordered pieces.
 ACCESSION BX004860
 VERSION BX004860.1 GI:26985650
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Dantio rerio (zebrafish)
 ORGANISM Dantio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 101080)
 AUTHORS Burton,J.
 TITLE Direct Submission

JOURNAL
 Submitted (10-DEC-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
 Clone requests: clonerequest@sanger.ac.uk
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zface@sanger.ac.uk
 ----- Project Information
 Center project name: dz144B24
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 99769 bases at least Q40
 Consensus quality: 100257 bases at least Q30
 Consensus quality: 100510 bases at least Q20
 Insert size: 100780; sum-of-contigs
 Insert size: 107197; 9.7% error; agarose-fp
 Quality coverage: 5.75x in Q20 bases; sum-of-contigs Quality
 coverage: 5.56x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 34070: contig of 34070 bp in length
 * 34071 34170: gap of 100 bp
 * 34171 67630: contig of 33460 bp in length
 * 67631 67730: gap of 100 bp
 * 67731 81714: contig of 13984 bp in length
 * 81715 81814: gap of 100 bp
 * 81815 101080: contig of 19266 bp in length.
 * Location/Qualifiers
 1..101080
 /organism="Dantio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="BUSM1-144B24"
 /clone_1b="BUSM1"
 1..34070
 /note="assembly_fragment:00417
 fragment_chain:1
 clone_end:77
 vector_side:left"
 34171..67630
 /note="assembly_fragment:00598
 fragment_chain:1"
 67731..81714
 /note="assembly_fragment:00262
 fragment_chain:1"
 81815..101080
 /note="assembly_fragment:00119
 fragment_chain:1
 clone_end:SP6
 vector_side:right"

ORIGIN
 Query Match 0.7%; Score 23; DB 2; Length 101080;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGAAACAAATCAACTGA 118
 |||||
 Db 39791 AATGAAACAAATCAACTGA 39769
 |||||

RESULT 12
 AL845510/c 101241 bp DNA linear VRT 30-JAN-2003
 LOCUS AL845510/c

DEFINITION Zebrafish DNA sequence from clone BUSM1-132M23 in linkage group 7, complete sequence.

ACCESSION AL845510 GI:28172206

VERSION AL845510.8

KEYWORDS HTG.

SOURCE Dantio rerio (zebrafish)

ORGANISM Dantio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Dantio.

AUTHORS Bates, K.

TITLE Direct Submission

JOURNAL Submitted (29-JAN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jan 30, 2003 this sequence version replaced gi:27801659.

COMMENT ----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Bm, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the NCBI database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

beginning 'dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dtr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml

BUSM1-132M23 is from a Zebrafish PAC library

VECTOR: pCYPAC-6.

FEATURES

source

1..101241

/organism="Dantio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="BUSM1-132M23"

/clone_1lb="BUSM1"

ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 101241;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGAAACAAATCAACTGA 118

DB 39751 AATGAAACAAATCAACTGA 39729

RESULT 13

AC107028/c 126716 bp DNA linear PRI 29-MAR-2003

LOCUS Homo sapiens 3 BAC RP11-547K2 (Roswell Park Cancer Institute Human BAC library) complete sequence.

DEFINITION BAC library complete sequence.

ACCESSION AC107028

VERSION AC107028.5 GI:28006935

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbarella, J., Benton, J., Bimge, K., Blankenburg, K., Bonnin, D., Buck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davis, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, M., Louisse, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mochab, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nicholson, E., Norken, S., Ogih, M., Okunoda, G., Oregu, N., Overduin, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojoubokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, P., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usami, K., Vazquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G., and Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (14-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (29-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 5 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 6 (bases 1 to 126716)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 7 (bases 1 to 126716)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 8 (bases 1 to 126716)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 9 (bases 1 to 126716)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jan 29, 2003 this sequence version replaced gi:20340431.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
 STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) For Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source Location/Qualifiers
 1..126716
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-547K2"
 complement(1..3011)
 /note="overlaps bases 1..3011 of clone ACO26305"
 /function="clone overlap"
 complement(28..209)
 repeat_region
 misc_feature
 repeat_region

repeat_region	/rpt_family="MER34"	219..973
repeat_region	/rpt_family="LM4"	1409..1455
repeat_region	/rpt_family="AluY"	1456..1754
repeat_region	/rpt_family="AluY"	1762..1998
repeat_region	/rpt_family="AluSc"	2145..2195
repeat_region	/rpt_family="AT_rich"	2612..2921
repeat_region	/rpt_family="AluSx"	3493..3520
repeat_region	/rpt_family="AT_rich"	3573..3598
repeat_region	/rpt_family="AT_rich"	5930..5955
repeat_region	/rpt_family="AT_rich"	complement(6168..6478)
repeat_region	/rpt_family="AluSp"	6945..6984
repeat_region	/rpt_family="(TTTA)n"	complement(6986..7276)
repeat_region	/rpt_family="AluSx"	11796..11975
repeat_region	/rpt_family="MER5A"	complement(13903..13967)
repeat_region	/rpt_family="L2"	14192..14235
repeat_region	/rpt_family="MIR"	16052..16078
repeat_region	/rpt_family="AT_rich"	complement(17082..17461)
repeat_region	/rpt_family="WSTD"	complement(17583..17934)
repeat_region	/rpt_family="MER21B"	17944..17997
repeat_region	/rpt_family="(TTATA)n"	complement(18003..18257)
repeat_region	/rpt_family="AluDb"	complement(18261..18557)
repeat_region	/rpt_family="MER21B"	19148..19215
repeat_region	/rpt_family="AT_rich"	19234..19342
repeat_region	/rpt_family="FLAM_C"	19377..19428
repeat_region	/rpt_family="A-rich"	

Query Match 0.7%; Score 23; DB 9; Length 126716;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47492 TCTTTACTCTTTAATTTTCT 1695
 1673 TCTTTACTCTTTAATTTTCT 1695

RESULT 14
 LOCUS BX005128/c
 DEFINITION zebrafish DNA sequence from clone CH211-129H4 in linkage group 17,
 complete sequence.
 ACCESSION BX005128
 VERSION BX005128.4 GI:37518236
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 152345)

AUTHORS
TITLE
JOURNAL
COMMENT

COMMENT

Sehra, H.
Direct Submission
Submitted (03-Oct-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 5, 2003 this sequence version replaced gi:30141705.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zf1sh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, SW, SWISSPROT, Tr, TREMBL, WP, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
beginning 'Dr' were identified by the Repeat discovery system
(Zhong Bao and Sean Eddy, submitted), and those beginning 'dir'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see
http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml
CH211-129H14 is from a CHORI-211 BAC library

VECTOR: pTRABAC2.1
Derived zebrafish pUC subclones occasionally display
inconsistency over the length of mononucleotide A/T runs and
conserved TA repeats. Where this is found the longest good quality
representation will be submitted.

FEATURES

source

1. 152345
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-129H14"
/clone_lib="CHORI-211"

ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 152345;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3141 TTAGTATTTGTTTTTTTTTGT 3153
DB 2054 TTAGTATTTGTTTTTTTTTGT 2032

RESULT 15
AC048381 180925 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 3 clone RP11-547K2 map 3, WORKING DRAFT
DEFINITION
SEQUENCE, 37 unordered pieces.
AC048381
AC048381.3 GI:8084282
VERSION HTG, HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 180925)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-547K2
Unpublished
2 (bases 1 to 180925)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Baettler, V., Bedalov, F.,
Boguski, M., Boulton, A., Brown, A., Burkett, G.,
Campomano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeRubeis, R., Dewar, K., Diaz, J.S.,
Dodge, S., Domini, M., Doyle, M., Ferrer, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hago, B., Heald, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,
Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lechoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R.,
Meldrum, J., Meneses, L., Minova, T., Miranda, C., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange, R., Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (14-Apr-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 180925)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Baettler, V., Bedalov, F.,
Boguski, M., Boulton, A., Brown, A., Burkett, G.,
Campomano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeRubeis, R., Dewar, K., Diaz, J.S.,
Dodge, S., Domini, M., Doyle, M., Ferrer, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hago, B., Heald, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,
Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lechoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R.,
Meldrum, J., Meneses, L., Minova, T., Miranda, C., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange, R., Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:1770638.
All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L9685

Center clone name: 547 K.2

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158029 bases at least Q40

Consensus quality: 173285 bases at least Q20

Insert size: 183000; agarose-IP

Insert size: 177325; sum-of-contrigs
Quality coverage: 3.4 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently consists of 37 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1010: contrig of 1010 bp in length
1011 1110: gap of 100 bp
1111 2487: contrig of 1377 bp in length
2488 2587: gap of 100 bp
2588 3720: contrig of 1133 bp in length
3721 3820: gap of 100 bp
3821 4892: contrig of 1072 bp in length
4893 4992: gap of 100 bp
4993 5489: contrig of 497 bp in length
5490 5589: gap of 100 bp
5590 7151: contrig of 1562 bp in length
7152 7251: gap of 100 bp
7252 8655: contrig of 1404 bp in length
8656 8755: gap of 100 bp
8756 10134: contrig of 1379 bp in length
10135 10234: gap of 100 bp
10235 13107: contrig of 2873 bp in length
13108 13207: gap of 100 bp
13208 15265: contrig of 2058 bp in length
15266 15365: gap of 100 bp
15366 17768: contrig of 2403 bp in length
17769 17868: gap of 100 bp
17869 20233: contrig of 2365 bp in length
20234 20333: gap of 100 bp
20334 22968: contrig of 2635 bp in length
22969 23068: gap of 100 bp
23069 25333: contrig of 2265 bp in length
25334 25433: gap of 100 bp
25434 28978: contrig of 3545 bp in length
28979 29078: gap of 100 bp
29079 31604: contrig of 2526 bp in length
31605 31704: gap of 100 bp
31705 34500: contrig of 2696 bp in length
34501 34600: gap of 100 bp
34601 37255: contrig of 2755 bp in length
37256 37355: gap of 100 bp
37356 40057: contrig of 2702 bp in length
40058 40157: gap of 100 bp
40158 44190: contrig of 4033 bp in length
44191 44290: gap of 100 bp
44291 46785: contrig of 2495 bp in length
46786 46885: gap of 100 bp
46886 52597: contrig of 5712 bp in length
52598 52697: gap of 100 bp
52698 59026: contrig of 6329 bp in length
59027 59126: gap of 100 bp
59127 64431: contrig of 5305 bp in length
64432 64531: gap of 100 bp
64532 69115: contrig of 4584 bp in length
69116 69215: gap of 100 bp
69216 73866: contrig of 4651 bp in length
73867 73966: gap of 100 bp
73967 79516: contrig of 5550 bp in length
79517 79616: gap of 100 bp
79617 85688: contrig of 6072 bp in length
85689 85788: gap of 100 bp
85789 92058: contrig of 6270 bp in length
92059 92158: gap of 100 bp
92159 99039: contrig of 6881 bp in length
99040 99139: gap of 100 bp
99140 108885: contrig of 9746 bp in length

```

```

* 108886 108985: gap of 100 bp.
* 108986 116971: contrig of 7986 bp in length
* 116972 117071: gap of 100 bp
* 117072 126718: contrig of 9647 bp in length
* 126719 126818: gap of 100 bp
* 126819 137354: contrig of 10536 bp in length
* 137355 137454: gap of 100 bp
* 137455 146943: contrig of 9489 bp in length
* 146944 147043: gap of 100 bp
* 147044 157785: contrig of 10742 bp in length
* 157786 157885: gap of 100 bp
* 157886 180925: contrig of 23040 bp in length.

```

FEATURES

```

source
1..180925
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-547K2"
/clone_11b="RP11 Human Male BAC"
1..1010
/note="assembly_fragment"
misc_feature
111..2487
/note="assembly_fragment"
misc_feature
2588..3720
/note="assembly_fragment"
misc_feature
3821..4892
/note="assembly_fragment"
misc_feature
4993..5489
/note="assembly_fragment"
clone_end:T7
vector_side:right"
5590..7151
/note="assembly_fragment"
misc_feature
7252..8655
/note="assembly_fragment"
misc_feature
8756..10134
/note="assembly_fragment"
misc_feature
10235..13107
/note="assembly_fragment"
misc_feature
13208..15265
/note="assembly_fragment"

```

```

Query Match 0.7%; Score 23; DB 2; Length 180925;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1673 TCTTTACTTCTTAATTTTCT 1695
Db 16692 TCTTTACTTCTTAATTTTCT 16714

```

Search completed: April 9, 2004, 05:51:24
Job time : 8467.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 16:25:39 ; Search time 837.022 Seconds
(without alignments)
16621.841 Million cell updates/sec

Title: US-10-030-829-1
Perfect score: 3275
Sequence: 1 gcaacaacaacaataa.....caaatataggtggtacaaat 3275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3275	100.0	3275	4	AAF25373 Genomic s
2	963	29.4	1878	4	AAF25374 Nucleotid
3	27	0.8	27	4	AAF25372 PCR prime
4	23	0.7	23	4	AAF25371 PCR prime
5	22	0.7	6136	6	ABK40041 Human che
6	22	0.7	6301	6	ABN80100 Human che
7	22	0.7	9888	6	ABL33241 Human imm
8	21	0.6	6048	6	ABL32509 Human imm
9	21	0.6	6072	6	ABL32031 Human imm
10	21	0.6	8245	4	AA546448 Tumour su
11	21	0.6	13453	6	ABN80176 Human che
12	21	0.6	52302	8	ADA02738 Human CCN
13	21	0.6	52302	9	AD872476 Human CCN
14	21	0.6	52302	9	ADC85218 Human Ccn
15	20	0.6	420	3	AA525596 Arabidops
16	20	0.6	421	3	AA54974 Arabidops
17	20	0.6	476	6	AB278073 Human bre
18	20	0.6	505	3	AA560012 Human sec
19	20	0.6	536	6	AB034916 Oligonuc1
20	20	0.6	536	6	AB034917 Oligonuc1
21	20	0.6	536	6	AAFI2773 Aspergill
22	20	0.6	627	6	AB052242 Oligonuc1
23	20	0.6	627	6	AB052243 Oligonuc1

ALIGNMENTS

C	24	20	0.6	654	2	AA27594 Nucleotid
C	25	20	0.6	700	4	AA193017 Human inf
C	26	20	0.6	764	6	AB030109 Oligonuc1
C	27	20	0.6	764	6	AB030108 Oligonuc1
C	28	20	0.6	843	6	ABK48300 Human inh
C	29	20	0.6	928	9	ADD47853 Rat gene
C	30	20	0.6	928	9	ADD47852 Rat gene
C	31	20	0.6	928	9	ADD48892 Rat gene
C	32	20	0.6	1023	6	ABT07654 Human bre
C	33	20	0.6	1106	4	AA125105 Human bre
C	34	20	0.6	1461	7	ACA23079 Prokaryot
C	35	20	0.6	1876	5	ABA20651 Human ner
C	36	20	0.6	2501	9	ADB54297 Pretreat
C	37	20	0.6	2501	9	ADB54169 Pretreat
C	38	20	0.6	3077	7	AB210070 Haematopo
C	39	20	0.6	3077	7	AB210216 Haematopo
C	40	20	0.6	5278	4	AA546376 Tumour su
C	41	20	0.6	5278	6	ABL32823 Human imm
C	42	20	0.6	5376	6	ABL34150 Human imm
C	43	20	0.6	5453	6	ABK31481 Signal tr
C	44	20	0.6	5453	6	ABL70450 Chemicall
C	45	20	0.6	5453	6	AA61413 Human gen

RESULT 1
AAF25373
ID AAF25373 strand; DNA; 3275 BP.
XX
AC AAF25373;
XX
DT 15-MAY-2001 (first entry)
XX
DE Genomic sequence of the Arabidopsis SGS3 gene.
XX
KM SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW viral resistance; resistance; fatty acid content; protein content; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT exon 696..1658
FT intron /*tag= a
FT exon 1659..1731
FT intron /*tag= b
FT exon 1732..2023
FT intron /*tag= c
FT intron 2024..2134
FT exon /*tag= d
FT intron 2135..2379
FT exon /*tag= e
FT intron 2380..2481
FT exon /*tag= f
FT intron 2482..2648
FT exon /*tag= g
FT intron 2649..2738
FT exon /*tag= h
FT intron 2739..2949
FT exon /*tag= i
XX
PN WO200105951-A2.
XX
PD 25-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-FR002052.
XX
PR 16-JUL-1999; 99FR-00009417.
XX
PA 26-JAN-2000; 2000FR-00001006.
XX
XX (AVET) AVENTIS CROSCIENCE SA.
XX (INRG) INST NAT RECH AGRONOMIQUE.

XX Beclin C, Elmayer T, Vaucheret H;
 XX MPI; 2001-159529/16.
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 XX
 XX Claim 1; Page 31-32; 36pp; French.
 XX
 XX The present sequence represents the genomic sequence of the Arabidopsis
 CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
 CC inactivation (degradation of RNA) and for resistance to viruses.
 CC Overexpression of SGS3 results in plants with increased resistance to
 CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by
 CC expressing antisense RNA, by mutation or by homologous recombination)
 CC increases the level of the transgene product. This product may e.g.
 CC impart resistance (to herbicide, insects or pathogens), alter contents of
 CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an
 CC immunoglobulin or interferon
 CC
 XX Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
 SQ
 Query Match 100.0%; Score 3275; DB 4; Length 3275;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GACAAACAAACAAATTAACCAATGTCATGTTGCTAGCATTAATTAATAGGAAACA 60
 1 GACAAACAAACAAATTAACCAATGTCATGTTGCTAGCATTAATTAATAGGAAACA 60
 61 TTAAGTTAGCAAAAGAAAAAGAAAAAGTACAAAAATCAAAATCAATGTAAT 120
 61 TTAAGTTAGCAAAAGAAAAAGAAAAAGTACAAAAATCAAAATCAATGTAAT 120
 61 TTAAGTTAGCAAAAGAAAAAGAAAAAGTACAAAAATCAAAATCAATGTAAT 120
 121 GAAAAATTGAGATCGAATCGAAAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180
 121 GAAAAATTGAGATCGAATCGAAAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180
 121 GAAAAATTGAGATCGAATCGAAAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180
 181 ATTGTCCTCTTCTGCTGATTAATTTCTCCGAGTCTGATCACTACTCTCAC 240
 181 ATTGTCCTCTTCTGCTGATTAATTTCTCCGAGTCTGATCACTACTCTCAC 240
 181 ATTGTCCTCTTCTGCTGATTAATTTCTCCGAGTCTGATCACTACTCTCAC 240
 241 TCTCCGCGCTTAACTTAAGTCTCCGCTGTTTACTGTAAGTTTTCGCTTAAG 300
 241 TCTCCGCGCTTAACTTAAGTCTCCGCTGTTTACTGTAAGTTTTCGCTTAAG 300
 241 TCTCCGCGCTTAACTTAAGTCTCCGCTGTTTACTGTAAGTTTTCGCTTAAG 300
 301 CTTCCGATGCGCTCAACGCGATCTGAGCTGATTTCTCTTTTCTTCGCTGAAAA 360
 301 CTTCCGATGCGCTCAACGCGATCTGAGCTGATTTCTCTTTTCTTCGCTGAAAA 360
 301 CTTCCGATGCGCTCAACGCGATCTGAGCTGATTTCTCTTTTCTTCGCTGAAAA 360
 361 ATTGCGCTTAAGTCTGATTTGCAAGTTTGTGTAGGTTACTTTTTCCTTAATA 420
 361 ATTGCGCTTAAGTCTGATTTGCAAGTTTGTGTAGGTTACTTTTTCCTTAATA 420
 361 ATTGCGCTTAAGTCTGATTTGCAAGTTTGTGTAGGTTACTTTTTCCTTAATA 420
 421 TTTTAATAGTCTTAAGTAACGATACGAGTTTGTGCTTACGTTTGTGCTTAC 480
 421 TTTTAATAGTCTTAAGTAACGATACGAGTTTGTGCTTACGTTTGTGCTTAC 480
 421 TTTTAATAGTCTTAAGTAACGATACGAGTTTGTGCTTACGTTTGTGCTTAC 480
 481 TTTACCGTTTATGCTGATCGAGATTTGACGTGTAATAAAATCCTTCGTTTGT 540
 481 TTTACCGTTTATGCTGATCGAGATTTGACGTGTAATAAAATCCTTCGTTTGT 540
 481 TTTACCGTTTATGCTGATCGAGATTTGACGTGTAATAAAATCCTTCGTTTGT 540
 541 TTTGTTCAATTAATCGATTAATCTTGTGCTTGTGCTTGTGTTTGTGAGCCT 600
 541 TTTGTTCAATTAATCGATTAATCTTGTGCTTGTGCTTGTGTTTGTGAGCCT 600
 541 TTTGTTCAATTAATCGATTAATCTTGTGCTTGTGCTTGTGTTTGTGAGCCT 600
 601 ATGCGTTTGTGCTTGTGTTTCACTTCACTTCACTTCACTTCACTTCACTTCA 660
 601 ATGCGTTTGTGCTTGTGTTTCACTTCACTTCACTTCACTTCACTTCACTTCA 660
 601 ATGCGTTTGTGCTTGTGTTTCACTTCACTTCACTTCACTTCACTTCACTTCA 660
 661 ACTGTGGGTTTCTTTGCTGCTAATAGTTTGAATAAATGATTCATGAGGCTGAT 720

DB 661 ACTGTGGGTTTCTTTGCTGCTAATAGTTTGAATAAATGATTCATGAGGCTGAT 720
 QY 721 CTAGGAAAGAAAGAGTTCAAGGTTGTTATAGGCTGATGAGTTGAAACGTTCAAGTT 780
 QY 721 CTAGGAAAGAAAGAGTTCAAGGTTGTTATAGGCTGATGAGTTGAAACGTTCAAGTT 780
 DB 721 CTAGGAAAGAAAGAGTTCAAGGTTGTTATAGGCTGATGAGTTGAAACGTTCAAGTT 780
 QY 781 TGGCAGGAGCAAGCTGCTTCTTCAAGATGATGAGAGAGAGTGGAGTCAATTCCA 840
 QY 781 TGGCAGGAGCAAGCTGCTTCTTCAAGATGATGAGAGAGAGTGGAGTCAATTCCA 840
 DB 781 TGGCAGGAGCAAGCTGCTTCTTCAAGATGATGAGAGAGAGTGGAGTCAATTCCA 840
 QY 841 AAGAGAAACAAACCAACGAAACCTTCTGAAAACTTGGGTTTCTCAAAATTCGA 900
 DB 841 AAGAGAAACAAACCAACGAAACCTTCTGAAAACTTGGGTTTCTCAAAATTCGA 900
 QY 841 AAGAGAAACAAACCAACGAAACCTTCTGAAAACTTGGGTTTCTCAAAATTCGA 900
 QY 901 ATCCTCTAGAGCTTGGGTTGTCAGCAGCAAGGAGAGTACAGATATCTGGAGAG 960
 DB 901 ATCCTCTAGAGCTTGGGTTGTCAGCAGCAAGGAGAGTACAGATATCTGGAGAG 960
 QY 901 ATCCTCTAGAGCTTGGGTTGTCAGCAGCAAGGAGAGTACAGATATCTGGAGAG 960
 QY 961 GAAACAAATGATCCGGAGAGGTTAAACGCAATGTCGAGGCAATTCAGCTTAATCTG 1020
 DB 961 GAAACAAATGATCCGGAGAGGTTAAACGCAATGTCGAGGCAATTCAGCTTAATCTG 1020
 QY 961 GAAACAAATGATCCGGAGAGGTTAAACGCAATGTCGAGGCAATTCAGCTTAATCTG 1020
 DB 1021 GTGCGGAGCAGCTTGAACAGAAATGATTAACAATTTGTGACACCCCACTGTAT 1080
 QY 1021 GTGCGGAGCAGCTTGAACAGAAATGATTAACAATTTGTGACACCCCACTGTAT 1080
 DB 1021 GTGCGGAGCAGCTTGAACAGAAATGATTAACAATTTGTGACACCCCACTGTAT 1080
 QY 1081 CTGCGCCTCTTGAAGAGAGATGAAATGCGAGCAAGAGAGTTCTGTGACGACA 1140
 DB 1081 CTGCGCCTCTTGAAGAGAGATGAAATGCGAGCAAGAGAGTTCTGTGACGACA 1140
 QY 1081 CTGCGCCTCTTGAAGAGAGATGAAATGCGAGCAAGAGAGTTCTGTGACGACA 1140
 DB 1141 CAGCTGTGCAAGATTTCTGACAGTGAAGATGATGATGATGATGATGATGATGAT 1200
 QY 1141 CAGCTGTGCAAGATTTCTGACAGTGAAGATGATGATGATGATGATGATGATGAT 1200
 DB 1141 CAGCTGTGCAAGATTTCTGACAGTGAAGATGATGATGATGATGATGATGATGAT 1200
 QY 1201 ATGATTCGATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1201 ATGATTCGATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 QY 1201 ATGATTCGATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1261 ATGAGATCAAAAGAGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 QY 1261 ATGAGATCAAAAGAGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1261 ATGAGATCAAAAGAGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 QY 1321 GCAAGTGAATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 DB 1321 GCAAGTGAATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 QY 1321 GCAAGTGAATGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 DB 1381 CAGCTGTGCAAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 QY 1381 CAGCTGTGCAAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1381 CAGCTGTGCAAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 QY 1441 ATGCGAGCAAAAGAGCTAGGAGTTAAGCTCATGAGAAATGAGCTGAAGTTAG 1500
 DB 1441 ATGCGAGCAAAAGAGCTAGGAGTTAAGCTCATGAGAAATGAGCTGAAGTTAG 1500
 QY 1441 ATGCGAGCAAAAGAGCTAGGAGTTAAGCTCATGAGAAATGAGCTGAAGTTAG 1500
 DB 1501 AAAAGATCTAAGATGAGAGGCGCATCTGATCTTGTGCTGAGATTAATGGCAGT 1560
 QY 1501 AAAAGATCTAAGATGAGAGGCGCATCTGATCTTGTGCTGAGATTAATGGCAGT 1560
 DB 1501 AAAAGATCTAAGATGAGAGGCGCATCTGATCTTGTGCTGAGATTAATGGCAGT 1560
 QY 1561 GAAAGGTTGGTGAAGATGAAGAAATGAAATGCTGAGCTCAATGATGATCA 1620
 DB 1561 GAAAGGTTGGTGAAGATGAAGAAATGAAATGCTGAGCTCAATGATGATCA 1620
 QY 1561 GAAAGGTTGGTGAAGATGAAGAAATGAAATGCTGAGCTCAATGATGATCA 1620
 DB 1621 TCAATGAATACAGTGAATGAGAGCAATTAAGAGATTAAGATTAAGATTAAG 1680
 QY 1621 TCAATGAATACAGTGAATGAGAGCAATTAAGAGATTAAGATTAAGATTAAG 1680
 DB 1621 TCAATGAATACAGTGAATGAGAGCAATTAAGAGATTAAGATTAAGATTAAG 1680
 QY 1681 TTTCTTAATTTTCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1740
 DB 1681 TTTCTTAATTTTCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1740
 QY 1681 TTTCTTAATTTTCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1740
 DB 1741 ATGCGCAACAAAGAGCTGAGAAATCTTGAACAAGATGAGGCTTAGACACCAT 1800
 QY 1741 ATGCGCAACAAAGAGCTGAGAAATCTTGAACAAGATGAGGCTTAGACACCAT 1800

```

Db      1741 ATGGGCAACCAAGCTGCTGGAATACTTGACAACTATGAGGCTCTTAGAGCAGCCCAT 1800
Qy      TCCTATGATCCACAGGGCCATCGTGGGATGAGTGTCTGATGTTTGAAGACAGTGCACCT 1860
Db      1801 TCCTATGATCCACAGGGCCATCGTGGGATGAGTGTCTGATGTTTGAAGACAGTGCACCT 1860
Qy      1801 TCCTATGATCCACAGGGCCATCGTGGGATGAGTGTCTGATGTTTGAAGACAGTGCACCT 1860
Db      1861 GGCATATTTGAGGCCCAAGCCCTCCACCGGAGTTAGTGAATGAGGCTTTAGATAGAAAT 1920
Qy      1861 GGCATATTTGAGGCCCAAGCCCTCCACCGGAGTTAGTGAATGAGGCTTTAGATAGAAAT 1920
Db      1861 GGCATATTTGAGGCCCAAGCCCTCCACCGGAGTTAGTGAATGAGGCTTTAGATAGAAAT 1920
Qy      1921 GGCCTGGGGTCAAGAGCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATGAGCTTCTT 1980
Db      1921 GGCCTGGGGTCAAGAGCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATGAGCTTCTT 1980
Qy      1921 GGCCTGGGGTCAAGAGCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATGAGCTTCTT 1980
Db      1981 GCAACGAACCAAGATGAGCATATTCATCAACACTCTCAAGGCTCTGCCCCCAAG 2040
Qy      1981 GCAACGAACCAAGATGAGCATATTCATCAACACTCTCAAGGCTCTGCCCCCAAG 2040
Db      2041 AATTGATATATGCTTTTATGTTTGTCAATGGAATTTAAAGTTTGTGGTCCGTCTTA 2100
Qy      2041 AATTGATATATGCTTTTATGTTTGTCAATGGAATTTAAAGTTTGTGGTCCGTCTTA 2100
Db      2101 TGCATCTGTTATGATATATCTATGATTCATTGCGCAAAACAAGCTGAATTCAGTTG 2160
Qy      2101 TGCATCTGTTATGATATATCTATGATTCATTGCGCAAAACAAGCTGAATTCAGTTG 2160
Db      2161 AATCATACCAAGATGAGTGTGTAAGGAGCTGAGGAGATCTCTGAGGCAATCAGCAG 2220
Qy      2161 AATCATACCAAGATGAGTGTGTAAGGAGCTGAGGAGATCTCTGAGGCAATCAGCAG 2220
Db      2221 CTGAATCTATTTAAGAACAGCTCTCAAAACAGAACAGCAGCAAGGCTTTGAGAA 2280
Qy      2221 CTGAATCTATTTAAGAACAGCTCTCAAAACAGAACAGCAGCAAGGCTTTGAGAA 2280
Db      2281 TCTCTGGAATTTATGAGCGAAGAGCTCGTAGAATCGAGAGATATCGATCGTAGA 2340
Qy      2281 TCTCTGGAATTTATGAGCGAAGAGCTCGTAGAATCGAGAGATATCGATCGTAGA 2340
Db      2341 CAGAGAACTAAGATGAGCATGAAAGAGGAGGATGATTTTCTAGAGAAAT 2400
Qy      2341 CAGAGAACTAAGATGAGCATGAAAGAGGAGGATGATTTTCTAGAGAAAT 2400
Db      2401 CACAAACTTGACATTTTGTATTAACCTAATGATTCATTTTGTATATATGTCACAA 2460
Qy      2401 CACAAACTTGACATTTTGTATTAACCTAATGATTCATTTTGTATATATGTCACAA 2460
Db      2461 AAAAAGCTGTGTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
Qy      2461 AAAAAGCTGTGTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
Db      2521 CAGATCCAGAAAGAAAGAACGCAAGGAGGAAATTTGAGATGTTTGCAGCAGCAGAA 2580
Qy      2521 CAGATCCAGAAAGAAAGAACGCAAGGAGGAAATTTGAGATGTTTGCAGCAGCAGAA 2580
Db      2581 CGTGCCAAAGTTTGTGGCAGCAGCAGCAGCAATTAATCCCTGAGATGAGATTCG 2640
Qy      2581 CGTGCCAAAGTTTGTGGCAGCAGCAGCAGCAATTAATCCCTGAGATGAGATTCG 2640
Db      2641 CGAAAGAGTATATGATTAATCAATATATCCCTGAGATGAGATTCG 2700
Qy      2641 CGAAAGAGTATATGATTAATCAATATATCCCTGAGATGAGATTCG 2700
Db      2701 AAGATTAATGATTAATCCGTTTGTATGATGATGATGATGATGATGATGATGATGAT 2760
Qy      2701 AAGATTAATGATTAATCCGTTTGTATGATGATGATGATGATGATGATGATGATGAT 2760
Db      2761 ATCGAGTTTCAAGAGAAAGATGAGAGTTTGTGGAAGAGAGAGAGATGATGATGAT 2820
Qy      2761 ATCGAGTTTCAAGAGAAAGATGAGAGTTTGTGGAAGAGAGAGAGATGATGATGAT 2820
Db      2821 GATCAAGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGATGATGATGATGAT 2880
Qy      2821 GATCAAGAGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGATGATGATGATGAT 2880

```

```

Qy      2881 GAGAAAGATTTGATGAGGCTTTGGAAACAGCTCATGTACAAAGATGCGCTTCAATGAA 2940
Db      2881 GAGAAAGATTTGATGAGGCTTTGGAAACAGCTCATGTACAAAGATGCGCTTCAATGAA 2940
Qy      2941 GATGATTTGAGCAAAAGTCTGTATCAACAAGACATGATGATTTCTTTGTTGCTTTT 3000
Db      2941 GATGATTTGAGCAAAAGTCTGTATCAACAAGACATGATGATTTCTTTGTTGCTTTT 3000
Qy      3001 GATGTCGGAAGATGAGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 3060
Db      3001 GATGTCGGAAGATGAGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 3060
Qy      3061 GATTATGATTTATTTATCTCCAAATTTTGTAGACGATCTTAAGAGATTAAGTTCTT 3120
Db      3061 GATTATGATTTATTTATCTCCAAATTTTGTAGACGATCTTAAGAGATTAAGTTCTT 3120
Qy      3121 GTGACTAAACCAAGTTTCTTATGATATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 3180
Db      3121 GTGACTAAACCAAGTTTCTTATGATATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 3180
Qy      3181 TTAGACATTTTACCAAGCTCAGAGTGAATCAGAAATGCGAAATCAATGATTTT 3240
Db      3181 TTAGACATTTTACCAAGCTCAGAGTGAATCAGAAATGCGAAATCAATGATTTT 3240
Qy      3241 AGAATTTTATCTCAAAATATATGCGTACAAAT 3275
Db      3241 AGAATTTTATCTCAAAATATATGCGTACAAAT 3275

RESULT 2
AAF25374
ID      AAF25374 standard; cDNA; 1878 BP.
XX
AC      AAF25374;
XX
DT      15-MAY-2001 (first entry)
XX
DE      Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
XX
KW      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content; ss.
XX
OS      Arabidopsis thaliana.
XX
FH      Key
FT      CDS      Location/Qualifiers
FT      1..1878
FT      /*tag= a
FT      /product= "SGS3"
XX
PN      MO200105951-A2.
XX
PD      25-JAN-2001.
XX
PF      13-JUL-2000; 2000WO-FR002052.
XX
PR      16-JUL-1999; 99FR-00009417.
XX
PR      26-JAN-2000; 2000FR-00001006.
XX
PA      (AVET ) AVENTIS CROPS/SCIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI      Beclin C, Elmayer T, Vaucheret H;
XX
DR      WPI; 2001-159529/16.
XX      P-PSDB; AAB31798.
XX
PT      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
PS      Claim 1; Page 32-35; 36pp; French.
XX

```

The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide. The SGS3 gene is essential for post-transcriptional inactivation of (degradation of RNA) and for resistance to viruses. Overexpression of SGS3 results in plants with increased resistance to viruses, while inactivation of SGS3 in transgenic plants (e.g. by expressing antisense RNA, by mutation or by homologous recombination) increases the level of the transgene product. This product may e.g. impart resistance to herbicide, insects or pathogens, alter contents of essential fatty acids or proteins, or is pharmaceutically active, e.g. an immunoglobulin or interferon

Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 29.4%; Score 963; DB 4; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 696 ATGAGTTCTAGGCTGTGCTCAATGCTTAAGAAAGAAAGCTTCAGGGTGTATAGGCT 755
Db 1 ATGAGTTCTAGGCTGTGCTCAATGCTTAAGAAAGAAAGCTTCAGGGTGTATAGGCT 60
Oy 756 GAGGTGAACAGTGTGTTCAAGTTTGGCAGGAGCAGACTGGCTTCTTCAAGATGAT 815
Db 61 GAGGTGAACAGTGTGTTCAAGTTTGGCAGGAGCAGACTGGCTTCTTCAAGATGAT 120
Oy 816 GGAGAGAGTGGAGTGTCAATTTCCAAAGAAAGAAAGAAAGCACTTCTGGA 875
Db 121 GGAGAGAGTGGAGTGTCAATTTCCAAAGAAAGAAAGAAAGCACTTCTGGA 180
Oy 876 AAAAAGTGGTGTTCAGAAATTCGATCTTGAAGCTTGGGTGTGTCAGCAGAAAG 935
Db 181 AAAAAGTGGTGTTCAGAAATTCGATCTTGAAGCTTGGGTGTGTCAGCAGAAAG 240
Oy 936 AGAGGTAGCAACGTAATCTGGAGAGAGAAAGATATCCGGAGAGGTAAACGCAATGT 995
Db 241 AGAGGTAGCAACGTAATCTGGAGAGAGAAAGATATCCGGAGAGGTAAACGCAATGT 300
Oy 996 CGGGGATTCAGAGTACATATCTGTGCGGAGCAGCGTGTGAGCAAGAAAGTATGATAC 1055
Db 301 CGGGGATTCAGAGTACATATCTGTGCGGAGCAGCGTGTGAGCAAGAAAGTATGATAC 360
Oy 1056 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTTGGAAGAGATGATGGCAG 1115
Db 361 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTTGGAAGAGATGATGGCAG 420
Oy 1116 GGAAGAGAGAGTCTGCTCAGCAGACAGCTGTGAGAGATTTCTGACGTGAGATGAT 1175
Db 421 GGAAGAGAGAGTCTGCTCAGCAGACAGCTGTGAGAGATTTCTGACGTGAGATGAT 480
Oy 1176 GTGATATATCTTCTGAGAGAGAAATGATCCGATGCTTGGATGATCTGATGACGAC 1235
Db 481 GTGATATATCTTCTGAGAGAGAAATGATCCGATGCTTGGATGATCTGATGACGAC 540
Oy 1236 CTTGCAAGTATGATGATGATCTGAGATGTGATCAAAAGAGCCATGATCAAGAAAGCAG 1295
Db 541 CTTGCAAGTATGATGATGATCTGAGATGTGATCAAAAGAGCCATGATCAAGAAAGCAG 600
Oy 1296 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGAT 1355
Db 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGAT 660
Oy 1356 GAACCAAGAGAGAGAGTGTGCTGAGCTTGTGAGAGAGAGAGAGAGTGTGATGATGAT 1415
Db 661 GAACCAAGAGAGAGAGTGTGCTGAGCTTGTGAGAGAGAGAGAGAGTGTGATGATGAT 720
Oy 1416 TATAACCTGCACCTCTTCTAGAGCTATGCGAGCAAAAGAGAGTGTAGAGAGTAAAGCTC 1475
Db 721 TATAACCTGCACCTCTTCTAGAGCTATGCGAGCAAAAGAGAGTGTAGAGAGTAAAGCTC 780
Oy 1476 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 1535
Db 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 840

```

```

Oy 1536 CTTGTGTGATGATTTATGGGACAGTGAAGGGTTTGGTGTGAGATGAAAGATTAATGAA 1595
Db 841 CTTGTGTGATGATTTATGGGACAGTGAAGGGTTTGGTGTGAGATGAAAGATTAATGAA 900
Oy 1596 ATTGTGCGCTCCATGATGATCATCATGATTAATGATGATGATGATGATGATGAT 1655
Db 901 ATTGTGCGCTCCATGATGATCATCATGATTAATGATGATGATGATGATGATGAT 960
Oy 1656 AAG 1658
Db 961 AAG 963

```

RESULT 3
AAAF25372/c
ID AAF25372 standard; DNA; 27 BP.

AAAF25372;

AC AAF25372;

XX 15-MAY-2001 (first entry)

DE PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.

XX SGS3 gene; post-transcriptional inactivation; RNA degradation;

KW viral resistance; resistance; fatty acid content; protein content;

XX PCR primer; ss.

XX Arabidopsis thaliana.

XX WO200105951-A2.

XX 25-JAN-2001.

XX 13-JUL-2000; 2000WO-FR002052.

XX 16-JUL-1999; 99FR-00009417.

XX 26-JAN-2000; 2000FR-00001006.

XX (AVERT) AVENTIS CROPS/SCIENCE SA.

XX (INRG) INST NAT RECH AGRONOMIQUE.

XX Beclin C, Elmayan T, Vaucheret H;

XX WPI; 2001-159529/16.

XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus

XX resistance in plants and, when inhibited, for increasing transgene

XX expression.

XX Example 1; Page 22; 36pp; French.

XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis

XX thaliana SGS3 polypeptide. The SGS3 gene is essential for post-

XX transcriptional inactivation (degradation of RNA) and for resistance to

XX viruses. Overexpression of SGS3 results in plants with increased

XX resistance to viruses, while inactivation of SGS3 in transgenic plants

XX (e.g. by expressing antisense RNA, by mutation or by homologous

XX recombination) increases the level of the transgene product. This product

XX may e.g. impart resistance (to herbicide, insects or pathogens), alter

XX contents of essential fatty acids or proteins, or is pharmaceutically

XX active, e.g. an immunoglobulin or interferon

XX Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;

XX Query Match 0.8%; Score 27; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.055; 0; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0;

Oy 2926 GGCCTTCAATGAAGATGATGAGAC 2952

Db 27 GGCCTTCAATGAAGATGATGAGAC 1


```

RESULT 4
AAF25371
ID AAF25371 standard; DNA; 23 BP.
XX
XX AAF25371;
AC
XX
XX 15-MAY-2001 (first entry)
DT
XX
XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
DE
XX
XX SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW viral resistance; resistance; fatty acid content; protein content;
KW PCR primer; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200105951-A2.
PN
XX
XX 25-JAN-2001.
PD
XX
XX 13-JUL-2000; 2000MO-FR002052.
PF
XX
XX 16-JUL-1999; 99FR-00009417.
PR 26-JAN-2000; 2000FR-00001006.
XX
XX (AVET ) AVENTIS CROSCIENCE SA.
PA (INRG ) INST NAT RECH AGRONOMICUE.
XX
XX Beglin C, Elmayer T, Vaucheret H;
PI
XX
XX WPI; 2001-159529/16.
DR
XX
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression.
XX
XX Example 1; Page 22; 36pp; French.
XX
XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
CC transcriptional inactivation (degradation of RNA) and for resistance to
CC viruses. Overexpression of SGS3 results in plants with increased
CC resistance to viruses, while inactivation of SGS3 in transgenic plants
CC (e.g. by expressing antisense RNA, by mutation or by homologous
CC recombination) increases the level of the transgene product. This product
CC may e.g. impart resistance (to herbicide, insects or pathogens), alter
CC contents of essential fatty acids or proteins, or is pharmaceutically
CC active, e.g. an immunoglobulin or interferon
XX
XX
SQ Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;
Query Match 0.7%; Score 23; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 693 AAAATGAGTTCTAGGCGTGTCC 715
DB 1 AAAATGAGTTCTAGGCGTGTCC 23

```

```

XX
XX Homo sapiens.
OS
XX
XX WO200202806-A2.
PN
XX
XX 10-JAN-2002.
PD
XX
XX 29-JUN-2001; 2001WO-EP007470.
PF
XX
XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX
XX WPI; 2002-154757/20.
DR
XX
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
XX Claim 1; SEQ ID NO 123; 24pp; English.
PS
XX
XX The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DYPD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (St) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 6136 BP; 1661 A; 69 C; 1415 G; 2991 T; 0 U; 0 Other;
Query Match 0.7%; Score 22; DB 6; Length 6136;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3145 TATTTGTTTTTTTGGTAA 3166
DB 2122 TATTTGTTTTTTTGGTAA 2143

```

```

RESULT 5
ABK40041
ID ABK40041 standard; DNA; 6136 BP.
XX
XX ABK40041;
AC
XX
XX 21-MAY-2002 (first entry)
DT
XX
XX Human chemically pretreated gene sequence #62 strand 1.
DE
XX
XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; DYPD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

```

```

RESULT 6
ABN80100
ID ABN80100 standard; DNA; 6301 BP.
XX
XX ABN80100;
AC
XX
XX 15-JUL-2002 (first entry)
DT
XX
XX Human chemically modified disease associated gene SEQ ID NO 117.
DE
XX
XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation.

```


XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX (EPiG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful for
 XX diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT Claim 1; SEQ ID NO 482; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;
 Query Match 0.6%; Score 21; DB 6; Length 6048;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 526 TTGGTTTTGGTTTTGTTT 546
 Db 5395 TTGGTTTTGGTTTTGTTT 5415
 RESULT 9
 ABL32031
 ID ABL32031 standard; DNA; 6072 BP.
 AC ABL32031;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX
 DE Human immune system associated gene SEQ ID NO: 4.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosinatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200200928-A2.
 PN
 XX
 PD 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPiG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful for
 XX

PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 XX Claim 1; SEQ ID NO 4; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 6072 BP; 1595 A; 155 C; 1384 G; 2938 T; 0 U; 0 Other;
 Query Match 0.6%; Score 21; DB 6; Length 6072;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3145 TATTTTGTGTTTTTTGTTAA 3165
 Db 3717 TATTTTGTGTTTTTTGTTAA 3737
 RESULT 10
 AAS46448
 ID AAS46448 standard; DNA; 8245 BP.
 XX
 XX AAS46448;
 AC
 XX
 XX 18-DEC-2001 (first entry)
 DT
 XX
 DE Tumour suppressor gene derived chemically modified sequence #170.
 XX
 XX Human; tumour suppressor gene; oncogene; antitumour; cytosinatic; cancer;
 KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; de.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200168912-A2.
 PN
 XX
 PD 20-SEP-2001.
 XX
 XX 15-MAR-2001; 2001WO-EP002955.
 PF
 XX 15-MAR-2000; 2000DE-01013847.
 PR 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPiG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2001-602752/68.
 DR
 XX
 XX Fragments of chemically modified genes associated with tumor suppressor
 PT genes and oncogenes, useful in designing primers and probes for analyzing
 PT diseases associated with cytosine methylation state e.g. cancer.
 PT
 XX Claim 1; SEQ ID NO 170; 27pp; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and oncogenes
 CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
 CC 500 are missing from the sequence listing) sequences (Ss) and sequences
 CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
 CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
 CC probes for detecting the cytosine methylation state and/or single

nucleotide polymorphisms and also to be used in an array for analysing diseases associated with CpG dinucleotides e.g. cancers and tumours. The probes can also be used in a method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases, by analysing cytosine methylations. The parameters may be compared to another set of genetic and/or epigenetic parameters, which are disadvantageous to patients. The present sequence is one of the 533 genomic sequences derived from tumour suppressor genes and oncogenes. Sequences with even numbered Seq ID numbers are the complementary sequence of the corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence is missing). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 8245 BP, 1978 A; 305 C; 2119 G; 3842 T; 0 U; 1 Other;

Query Match 0.6%; Score 21; DB 4; Length 8245;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3141 TTAGTATTTTGTGTTTTTTTG 3161
|||||
6632 TTAGTATTTTGTGTTTTTTTG 6652

Db

RESULT 11
ABN80176
ID ABN80176 standard; DNA; 13453 BP.
AC ABN80176;
XX
XX
XX 15-JUL-2002 (first entry)
DT
DE
DE Human chemically modified disease associated gene SEQ ID NO 193.
XX
XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
XX antidiabetic; cytostatic; anticonvulsant; ds.
XX
XX Homo sapiens.
OS Synthetic.
CS
PN WO200200927-A2.
XX
XX 03-JAN-2002.
PD
XX 02-JUL-2001; 2001WO-EP007536.
PF
XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130908/17.
DR
XX Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
XX Claim 1; SEQ ID NO 193; 27BP; English.
PS
XX The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADPN, or APDI and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is

cc useful for the diagnosis or therapy of diseases associated with
cc development genes, in particular disease related to homeobox containing
cc gene (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
cc associated with congenital heart disease, epilepsy, diseases related to
cc histone deacetylation, Curranio syndrome, diseases with the
cc development of the brain and limb girdle muscular dystrophy and dwarfism.
cc Oligomers specific to each of the genes are useful for detecting the
cc methylation state of all CpG dinucleotides within the 350 sequences or
cc (ii) and their complementary sequences, (iii) and/or their complements and
cc the amplification of the 350 sequences, (iii) and/or their complements and
cc as oligomer probes for detecting the cytosine methylation state and/or
cc single nucleotide polymorphisms (SNPs). Note: The sequence data for this
cc patent did not form part of the printed specification but is based on
cc sequence information supplied to Derwent by the European Patent Office

xx
xx
SQ Sequence 13453 BP; 2569 A; 384 C; 3712 G; 6788 T; 0 U; 0 Other;

Query Match 0.6%; Score 21; DB 6; Length 13453;
Best Local Similarity 100.0%; Pred. NO. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3141 TTAGTATTGTTTGTGTTTGG 3161
|||||
Db 7909 TTAGTATTGTTTGTGTTTGG 7929

RESULT 12
ADA02738/C
ID ADA02738 standard; DNA; 52302 BP.
XX
XX AC
XX ADA02738;
XX
XX 06-NOV-2003 (first entry)
XX
XX Human CCND2 carcinoma associated gene, SEQ ID NO:1256.
DE
XX
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX gene; ds.
XX
XX Homo sapiens.
OS
XX
XX W02003057146-A2.
XX
XX 17-JUL-2003.
XX
XX 26-DEC-2002; 2002WO-US041414.
XX
XX 26-DEC-2001; 2001US-00035832.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX
XX Morris DW;
PI
XX
XX WPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1256; 245pp; English.

cc The invention relates to recombinant carcinoma associated (CA) nucleic
cc acid sequences from mouse and human (ADA01482-ADA03094), and to
cc recombinant carcinoma associated proteins (CAP) encoded by them. The
cc invention also encompasses expression vectors and host cells comprising a
cc CA nucleic acid, a polypeptide (especially an antibody) that specifically
cc binds to the protein, and a biochip comprising CA nucleic acid or
cc oncogenic retroviruses, which insert into the genome of the host organism
cc at random. Many of these do not carry transduced host oncogenes or
cc pathogenic trans-acting viral genes, meaning that cancer incidence is a
cc direct consequence of the effects of proviral integration into host
cc protooncogenes. The CA nucleic acid sequences can be used to diagnose

CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 8; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGCTTTGG 3001

DB 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 13
ADB72476/c
ID ADB72476 standard; DNA; 52302 BP.

XX ADB72476;

XX 04-DEC-2003 (first entry)

XX Human CCND2 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX Claim 1; SEQ ID NO 304; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human gene of the invention.

XX SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGCTTTGG 3001

DB 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 14
ADC85218/c
ID ADC85218 standard; DNA; 52302 BP.

XX ADC85218;

XX 01-JAN-2004 (first entry)

XX Human CCnd2 genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
XX secreted; transmembrane; intracellular; ds.

XX Homo sapiens.

XX WO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX The invention relates to a recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the fully defined carcinoma-
XX associated (CA) genes from the 50 tables given in the specification. The
XX CA proteins are secreted, transmembrane or intracellular proteins. The
XX recombinant nucleic acids are useful for screening for drug candidates
XX for diagnosing or treating carcinomas. Sequences given in ADC85215-
XX ADC85514 represent CA genes of the invention.

XX Claim 1; SEQ ID NO 4; 983pp; English.

XX SQ Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2981 GTTCTTTGTTTGCTTTGG 3001

DB 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 15
AAC52596/c
ID AAC52596 standard; DNA; 420 BP.

XX AAC52596;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 71849.

XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.

XX Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-00301439.
 XX
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136392P.
 PR 27-MAY-1999; 99US-0136782P.
 PR 28-MAY-1999; 99US-0137222P.
 PR 01-JUN-1999; 99US-0137528P.
 PR 03-JUN-1999; 99US-0137502P.
 PR 04-JUN-1999; 99US-0137724P.
 PR 07-JUN-1999; 99US-0138094P.
 PR 08-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 10-JUN-1999; 99US-0139119P.
 PR 14-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 16-JUN-1999; 99US-0139492P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 16-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144684P.
 PR 20-JUL-1999; 99US-0144684P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145224P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 28-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145919P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 05-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147302P.
 PR 06-AUG-1999; 99US-0147302P.
 PR 06-AUG-1999; 99US-0147302P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148171P.
 PR 12-AUG-1999; 99US-0148319P.
 PR 13-AUG-1999; 99US-0148319P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149168P.
 PR 16-AUG-1999; 99US-0149175P.
 PR 17-AUG-1999; 99US-0149426P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149829P.
 PR 20-AUG-1999; 99US-0149829P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0150566P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151303P.
 PR 01-SEP-1999; 99US-0151438P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161820P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 0.6%; Score 20; DB 3; Length 420;
Best Local Similarity 100.0%; Pred.No.11e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACAACCAACAAATTAAG 21
Db 378 ACAACCAACAAATTAAG 359

Search completed: April 9, 2004, 02:09:03
Job time : 842.022 secs

This Page Blank (uspto)

AU236368
 LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002
 DEFINITION AU236368 RAF1.4 Arabidopsis thaliana cDNA clone RAF1.4-93-K05 5', mRNA sequence.
 ACCESSION AU236368
 VERSION AU236368.1 GI:19875537
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 647)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@rc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plan/index_e.html) for further details.

FEATURES

source

Location/Qualifiers
 1..647
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAF1.4-93-K05"
 /issue_type="root"
 /lab_host="DH10B"
 /clone_lib="RAF1.4"
 /note="Site_1: BamHI; Site_2: SalI"

ORIGIN

Query Match 15.9%; Score 521; DB 9; Length 647;
 Best Local Similarity 100.0%; Pred. No. 1.1e-249;
 Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

687 GTTGTAAATGAGTCTAGAGGCTGCTCAATGCTTAGGAAAGAAACGTTGAGGGTGGT 746
 |||||
 83 GTTGTAAATGAGTCTAGAGGCTGCTCAATGCTTAGGAAAGAAACGTTGAGGGTGGT 142
 |||||
 747 TATAGGCTGAGGTTGACAGTGTGTTCAAGGTTTGGCAGGAGCAGACTGGCTTCTCA 806
 |||||
 143 TATAGGCTGAGGTTGACAGTGTGTTCAAGGTTTGGCAGGAGCAGACTGGCTTCTCA 202
 |||||
 807 CAAAGATGAGGAGAGTGGAGGTCATTTCGAAGAACAAGAACCAACAGAGAAC 866
 |||||
 203 CAAAGATGAGGAGAGTGGAGGTCATTTCGAAGAACAAGAACCAACAGAGAAC 262
 |||||
 867 ACTTGTGAAAACTGGGTTTCTCAGAAATGGAATCTCTGAGCTTGGGGTGTGACG 926
 |||||
 263 ACTTGTGAAAACTGGGTTTCTCAGAAATGGAATCTCTGAGCTTGGGGTGTGACG 322
 |||||
 927 CAGCAAGGAGAGTNGACAAGTATCTGGAGAGAGAAACAATGTATCCGGAGAGTAAAC 986
 |||||
 323 CAGCAAGGAGAGTNGACAAGTATCTGGAGAGAGAAACAATGTATCCGGAGAGTAAAC 382
 |||||
 987 GCGAATGCTCGGGGCACTTCAAGCTAACATATCTGCTCGGGAGCAGACCTTGAACAGAAAG 1046
 |||||
 383 GCGAATGCTCGGGGCACTTCAAGCTAACATATCTGCTCGGGAGCAGACCTTGAACAGAAAG 442

QY 1047 TATGATPAACACTTGTGTGACACCCCACTGTATCTCGCCCTCTTGGAGAGAGATGG 1106
 |||||
 DB 443 TATGATPAACACTTGTGTGACACCCCACTGTATCTCGCCCTCTTGGAGAGAGATGG 502
 |||||
 QY 1107 AATTGGCAGGCAAGAGAGAGTCTGTCTGACACACAGCTGTGACAGAGTTTCTTGAACGTG 1166
 |||||
 DB 503 AATTGGCAGGCAAGAGAGAGTCTGTCTGACACACAGCTGTGACAGAGTTTCTTGAACGTG 562
 |||||
 QY 1167 GAGGATGATGTGATTAATCTCTGAGAGAGAGATGATTC 1207
 |||||
 DB 563 GAGGATGATGTGATTAATCTCTGAGAGAGAGATGATTC 603
 |||||

RESULT 4
 LOCUS AQ964580/c 613 bp DNA linear GSS 28-JUN-2000
 DEFINITION LERGX20TF LERG Arabidopsis thaliana genomic clone LERGX20, genomic survey sequence.
 ACCESSION AQ964580
 VERSION AQ964580.1 GI:6792281
 KEYWORDS GSS

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 613)
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T., Feldblyum, T., Liang, P., Creasy, T., and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atel@igr.org
 For additional information, see <http://www.tigr.org/tdb/ac/at.html>
 Seq primer: TP
 Class: shotgun.

FEATURES
 source Location/Qualifiers

1..613
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERRECTA"
 /db_xref="taxon:3702"
 /clone="LERGX20"
 /clone_lib="LERGX"
 /note="Torgan: Leaf; Vector: pUC19UR; Total genomic DNA was sheared to 0.4-0.7 Kbp before ligation."

ORIGIN

Query Match 15.4%; Score 504; DB 28; Length 613;
 Best Local Similarity 99.7%; Pred. No. 3.8e-241;
 Matches 604; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

615 TGTATTAACCTTCAAGCTTCAATGTTGAGATTTTGGTAGTGAAGTGGGTTTCTT 674
 |||||
 DB 606 TGTATTAACCTTCAAGCTTCAATGTTGAGATTTTGGTAGTGAAGTGGGTTTCTT 547
 |||||
 675 TGTGTGCTATAGTGTGTAATAATGAGTCTAGGGCTGCTCAATGCTTAAGAAAGAAC 734
 |||||
 DB 606 TGTGTGCTATAGTGTGTAATAATGAGTCTAGGGCTGCTCAATGCTTAAGAAAGAAC 487
 |||||
 546 TGTGTGCTATAGTGTGTAATAATGAGTCTAGGGCTGCTCAATGCTTAAGAAAGAAC 487
 |||||
 735 GTTCAGGCTGTATATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGGA 794
 |||||
 DB 486 GTTCAGGCTGTATATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGGA 427
 |||||
 795 CTGGCTTTCTTCAAGATGATGAGAGAGTGGAGGTCATTTCCAGAAAGAACAGAAC 854
 |||||
 DB 426 CTGGCTTTCTTCAAGATGATGAGAGAGTGGAGGTCATTTCCAGAAAGAACAGAAC 367
 |||||

QY	855	AAACACAGAAAACACTTCTGGAAAAAAGTTGGTTCTCAGAAATTCGAATCCCTCAGAGCT	914
Db	366	AAATCAGGAAACACTTCTGGAAAAAAGTTGGTTCTCAGAAATTCGAATCCCTCAGAGCT	307
QY	915	TGGGGTGGTCAGCAGCAAGGAGAGTAGCAACGATCTGGGAGAGGAAACAATGTATCC	974
Db	306	TGGGGTGGTCAGCAGCAAGGAGAGGTATGCAACGATCTGGGAGAGGAAACAATGTATCC	247
QY	975	GGGAGAGGTAAACGGCAATGTTGGGGGCAATTCAAGTAAACATTTGTTGGGGAGAGAGCG	1034
Db	246	GGGAGAGGTAAACGGCAATGTTGGGGGCAATTCAAGTAAACATTTGTTGGGGAGAGAGCG	187
QY	1035	TTGAGCAGAAAGTATGATTAACAATTGTTGGCACCACCCTGATATCGGCGCTCCTTTG	1094
Db	186	TTGAGCAGAAAGTATGATTAACAATTGTTGGCACCACCCTGATATCGGCGCTCCTTTG	127
QY	1095	GAAAGAGATGAAATTGGCAGGCAAGAGAGGTTCTGTCAGCACACGTTGTGAGAG	1154
Db	126	GAAAGAGATGAAATTGGCAGGCAAGAGAGGTTCTGTCAGCACACGTTGTGAGAG	67
QY	1155	TTTCTCTGACGTGAGAGATGATGTGTGATTAATGCTTCTAGGAAAGAAATTCGATGCT	1214
Db	66	TTTCTCTGACGTGAGAGATGATGTGTGATTAATGCTTCTAGGAAAGAAATTCGATGCT	7
QY	1215	TTGGAT 1220	
Db	6	TTGGAT 1	

Qy	1742	TGGGCAACCAAGAGCTCTGGAATATCTTCACAAGATGTAGAGCTCTTAGAGCAGGCATT	1801
Db	67	TGGGCAACCAAGAGCTCTGGAATATCTTCACAAGATGTAGAGCTCTTAGAGCAGGCATT	126
Qy	1802	CCTATGGTCCACAGGGCCATCGTGGAGTGAATGATGTTCTGATGTTTGAAGCAGTCCACTG	1861
Db	127	CCTATGGTCCACAGGGCCATCGTGGAGTGAATGATGTTCTGATGTTTGAAGCAGTCCACTG	186
Qy	1862	GCTATTTTGGAGGCCGAACGGCTCCACCCGGAGTTAGCTGAGATGGGGTTAGATGAATTG	1921
Db	187	GCTATTTTGGAGGCCGAACGGCTCCACCCGGAGTTAGCTGAGATGGGGTTAGATGAATTG	246
Qy	1922	CCTGGGGTCAAGACGCGATGTTTCTCGAGGTGTTCCGCACTGATGGCTTCCTTG	1981
Db	247	CCTGGGGTCAAGACGCGATGTTTCTCGAGGTGTTCCGCACTGATGGCTTCCTTG	306
Qy	1982	CAACGAAGCAAGATCTGGAACATATTCATCAACTCTCAAGTTCCTCCCCCAAGAA	2041
Db	307	CAACGAAGCAAGATCTGGAACATATTCATCAACTCTCAAGTTCCTCCCCCAAGAA	366
Qy	2042	ATTGGAATATGCTTTAGTTTGTCAATGGAATTTAAAGTTTGGTGCAGTTAAT	2101
Db	367	ATTGGAATATGCTTTAGTTTGTCAATGGAATTTAAAGTTTGGTGCAGTTAAT	426
Qy	2102	GCATCTGTAATGATATATCTATGATTCATT	2132
Db	427	GCATCTGTAATGATATATCTATGATTCATT	457

RESULT 5				
CNS009PH				
LOCUS	CNS009PH	457 bp	DNA	linear GSS 28-JUN-1999
DEFINITION	Arabidopsis thaliana genome survey sequence T7 end of BAC F6G21 of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.			
ACCESSION	AL084227			
VERSION	AI084227.1	GI:5285367		
KEYWORDS	GSS.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Angiosperms; Magnoliopsida; eudicotyledons; core eudicots; Spermatophytes; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 457) Salanoubat,M., Choise,N., Artiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 457)			
JOURNAL	Genoscope.			
REFERENCE	Direct Submission			
TITLE	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :			
JOURNAL				

```

FEATURES
    source
        Location/Qualifiers
            1..457
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone="F8G21"
                /clone_1fb="1Gf"
                /note="end : T7"

```

ORIGIN

Query Match	13.8%	Score 451	DB 29	length 457
Best Local Similarity	100.0%	Pred. No.	1.6e-214	
Matches 451; Conservative	0	Mismatches	0	Indels 0
				Gaps 0

QY 1682 TCTTAAATTTTCCTGCAATCTACTGATTAGAAATGTACATTTGAGGGCTCGGA 174
Db 7 TCTTAATTTTCTCTGCATCTACTGATCTTAGAATGTACATTTGAGGCTCGGCA 66

FEATURES	Location/Qualifiers
source	1. .421

```

FEATURES
    source
        location/Qualifiers
            1..421
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /db_xref="taxon:3702"
            /clone="RAFL14-93-K05"
            /tissue_type="root"
            /lab_host="RDH10B"
            /clone_lib="RAFL14"
            /note="Site_1: BamHI; Site_2: SalI"

```


/strain="Columbia"
/db_xref="taxon:3702"
/clone="APD25402R"
/issue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 10.0%; Score 329; DB 9; Length 380;
Best Local Similarity 99.7%; Pred. No. 3.1e-153;
Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 971 ATCCGGAGAGGTAACGCAATGTCGGGCACTTCAAGTAACTATCTGGTCCGGAGCG 1030
DB 1 ATCCGGAGAGGTAACGCAATGTCGGGCACTTCAAGTAACTATCTGGTCCGGAGCG 60
QY 1031 AGCGTTGAGCAAAATGATGATTAACAATTGTCGACCCCACTGATCTCGCCCTCC 1090
DB 61 AGCGTTGAGCAAAATGATGATTAACAATTGTCGACCCCACTGATCTCGCCCTCC 120
QY 1091 TTGGAGAGAGATGATGATGAGGCAAGAGAGGTTCTGCTCAGCAGACAGCTGTGCA 1150
DB 121 TTGGAGAGAGATGATGATGAGGCAAGAGAGGTTCTGCTCAGCAGACAGCTGTGCA 180
QY 1151 GGAATTTCTGACGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1210
DB 181 GGAATTTCTGACGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 1211 TGCTTGGATGATCTGATGACGACCTTGCAGATGATGATGATGATGATGATGATGAT 1270
DB 241 TGCTTGGATGATCTGATGACGACCTTGCAGATGATGATGATGATGATGATGATGAT 300
QY 1271 AAAGAGCCATGATCAGAAAGCAATTAAGTGTTCAAAAGTCTTGGCAGCTTGA 1330
DB 301 AAAGAGCCATGATCAGAAAGCAATTAAGTGTTCAAAAGTCTTGGCAGCTTGA 360
QY 1331 TAGCTTTCGATCCGAGCAGA 1350
DB 361 TAGCTTTCGATCCGAGCAGA 380

RESULT 9
A0010650 290 bp DNA linear GSS 29-MAY-1998
LOCUS F27C8TRC IGF Arabidopsis thaliana genomic clone F27C8, genomic
DEFINITION survey sequence.
ACCESSION A0010650
VERSION A0010650.1 GI:3165927
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 290)
Rounsley,S.D., Suh,E.J., Wible,C., Golden,K., Shatsman,S., Choi,P.,
Yu,K., Akintetoye,B., Shen,K., Goomasekaram,S., Miltchev,J.,
Adams,M.D. and Venter,J.C.
A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 4

JOURNAL
COMMENT Unpublished (1998)
Other GSSs: F27C8TRC
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse

Class: BAC ends
High quality sequence stop: 290.
Location/Qualifiers
1..290

FEATURES

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F27C8"
/sex="hermaphrodite"
/clone_lib="IGF"
/note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN

Query Match 8.9%; Score 290; DB 28; Length 290;
Best Local Similarity 100.0%; Pred. No. 1.3e-133;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1668 TTCTGCTTTTACTTCTTTAATTTTCTCTTGATTTCTACTGATTTGAAATGTTACTT 1727
DB 1 TTCTGCTTTTACTTCTTTAATTTTCTCTTGATTTCTACTGATTTGAAATGTTACTT 60
QY 1728 GTAGTGGCTGGCATGGGCAACCAAGAGCTGCTGAAATCTTGCAGCAATAGAGCTCT 1787
DB 61 GTAGTGGCTGGCATGGGCAACCAAGAGCTGCTGAAATCTTGCAGCAATAGAGCTCT 120
QY 1788 TAGAGCAGCCATTCTCATATGTCACAGGGCCATCGTGGATGATGTTCTGATGTTGA 1847
DB 121 TAGAGCAGCCATTCTCATATGTCACAGGGCCATCGTGGATGATGTTCTGATGTTGA 180
QY 1848 GAGCAGTCCACATGCTATTTTGAAGCCCAACCCCTCCACCGGAGTATGATGATGG 1907
DB 181 GAGCAGTCCACATGCTATTTTGAAGCCCAACCCCTCCACCGGAGTATGATGATGG 240
QY 1908 GTTAGATGAATTTGCTGGGGTCAAGCGCAGTATGTTTCTGAGGTG 1957
DB 241 GTTAGATGAATTTGCTGGGGTCAAGCGCAGTATGTTTCTGAGGTG 290

RESULT 10
A0959659 523 bp DNA linear GSS 28-JAN-2000
LOCUS LEREM21TR LERB Arabidopsis thaliana genomic clone LEREM21, genomic
DEFINITION survey sequence.
ACCESSION A0959659
VERSION A0959659.1 GI:6787360
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 523)
Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Uteerbach,T.,
Peldiblum,T., Liang,F., Creasy,T. and Frazer,C.M.
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based

JOURNAL
COMMENT Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: ac@tigr.org
For additional information, see <http://www.tigr.org/cdb/at/at.html>
Seq primer: TR
Class: shotgun.

FEATURES
source location/Qualifiers
1..523
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

/strain="LANDSBERG_ERECTA"
 /db_xref="taxon:3702"
 /clone="LBEREM21"
 /clone_lib="LBERE"
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was
 sheared to 0.6-0.8 Kbp before ligation."

ORIGIN

Query Match 8.3%; Score 273; DB 28; Length 523;
 Best Local Similarity 99.3%; Pred. No. 3.8e-125;
 Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 862 GAAACACTTCTGAAAACTTGGGTTCTCAGAAATTCAGATCCTCTAGAGCTTGGGGTG 921
 DB 25 GAAACACTTCTGAAAACTTGGGTTCTCAGAAATTCAGATCCTCTAGAGCTTGGGGTG 84
 QY 922 GTCAGCAGCAAGGAGAGGTAGCAAGTATCTGGAGAGAAACAATGTATCCGGAGAG 981
 DB 85 GTCAGCAGCAAGGAGAGGTAGCAAGTATCTGGAGAGAAACAATGTATCCGGAGAG 144
 QY 982 GTPAAGCGCAATGGTGGGCAATTCAGCTAACATCTGGTGGGAGCAAGCGTTGAGCA 1041
 DB 145 GTPAAGCGCAATGGTGGGCAATTCAGCTAACATCTGGTGGGAGCAAGCGTTGAGCA 204
 QY 1042 GAAAGTATGATACAACTTTGTGGCAACCCCACTGTATCTCGCCTCTTGGAGAGAG 1101
 DB 205 GAAAGTATGATACAACTTTGTGGCAACCCCACTGTATCTCGCCTCTTGGAGAGAG 264
 QY 1102 GATGAAATTTGGCAGGAGAGAGGTTCTGCTCAGACACAGCTGTGAGAGATTTCTTG 1161
 DB 265 GATGAAATTTGGCAGGAGAGAGGTTCTGCTCAGACACAGATGTGAGAGATTTCTTG 324
 QY 1162 AGCTGAGATGATATGATGATTAATGCTTCAGAGAAAGATGATTCGATGCTTGGAGAG 1221
 DB 325 AGCTGAGATGATATGATGATTAATGCTTCAGAGAAAGATGATTCGATGCTTGGAGAG 384
 QY 1222 ATTCTGATGACGACCTTGGCAAGTATGATGATGATGATGATGATGATGATGATGAT 1281
 DB 385 ATTCTGATGACGACCTTGGCAAGTATGATGATGATGATGATGATGATGATGATGATGAT 444
 QY 1282 GATCAC 1287
 DB 445 GATCAC 450

RESULT 11
 AL945644/c 536 bp DNA linear GSS 30-NOV-2002
 LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-290H04-015353,
 DEFINITION genomic survey sequence.
 ACCESSION AL945644
 VERSION AL945644.1 GI:24402266
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 Strizhov, N., Li, Y., Rosso, M., Vlehoever, P., Dekker, K., Saedler, H.
 and Weishaar, B.
 TITLE A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 JOURNAL Unpublished
 AUTHORS 2
 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weishaar, B.
 REFERENCE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 JOURNAL Unpublished
 AUTHORS 3 (bases 1 to 536)
 Li, Y., Rosso, M., Strizhov, N. and Weishaar, B.
 TITLE Direct Submission

JOURNAL

Submitted (21-OCT-2002) Weishaar B., Max-Planck-Institut fuer
 Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 COMMENT This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone MQM1. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
 http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES

source

Location/Qualifiers
 1..536

/organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-290H04-015353"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector pAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed."

ORIGIN

Query Match 7.1%; Score 232; DB 29; Length 536;
 Best Local Similarity 100.0%; Pred. No. 1.4e-104;
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3044 TGGGCAATCTAAGAGATTAATGATTAATTCCTCCAAATTTTATGAGAGGATCTAA 3103
 DB 527 TGGGCAATCTAAGAGATTAATGATTAATTCCTCCAAATTTTATGAGAGGATCTAA 468
 QY 3104 GGAAGCATTAAGTTCTTGTGATTAACCAAGTTTCTTATGATTTTGTGTTTGGT 3163
 DB 467 GGAAGCATTAAGTTCTTGTGATTAACCAAGTTTCTTATGATTTTGTGTTTGGT 408
 QY 3164 AAAATTCTATGAAAGTTAGACATATTACCAAGCTCAGAGTGAATCAGAAATGGCAA 3223
 DB 407 AAAATTCTATGAAAGTTAGACATATTACCAAGCTCAGAGTGAATCAGAAATGGCAA 348
 QY 3224 ATCAAAATCATGTTTATGAAATTTTATATCTCAAAATTTATGGGTACAAAT 3275
 DB 347 ATCAAAATCATGTTTATGAAATTTTATATCTCAAAATTTATGGGTACAAAT 296

RESULT 12
 AV566465/c 619 bp mRNA linear EST 07-SEP-2000
 LOCUS AV566465 Arabidopsis thaliana green siliques Columbia Arabidopsis
 DEFINITION thaliana cDNA clone SQ244b06f 3', mRNA sequence.
 ACCESSION AV566465
 VERSION AV566465.1 GI:8737917
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 Asamizu, E., Nakamura, Y., Seto, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7 (3), 175-180 (2000)
 MEDLINE 20363093
 PUBMED 10907847
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute

Yana 1532-3, Kiserazu, Chiba 292-0812, Japan
 Email: asami@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers

1. 619
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ24406F"
 /issue_type="green siliques"
 /clone_1b="Arabidopsis thaliana green siliques Columbia"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 6.4%; Score 210; DB 9; Length 619;
 Best Local Similarity 100.0%; Pred. No. 1.6e-93;
 Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2170 CAAGAGATGGTTGAAGAGAGCTGAGCAGATCTCTGAGACATACAGAGCTGAATAC 2229
 DB 619 CAGAGATGTTGAAGAGAGCTGAGCAGATCTCTGAGACATACAGAGCTGAATAC 560
 QY 2230 TTTAAGACCAAGCTCTCAAAACAGACAGACAGAGCTTGAAGATCTCTGAA 2289
 DB 559 TTTAAGACCAAGCTCTCAAAACAGACAGACAGAGCTTGAAGATCTCTGAA 500
 QY 2290 ATTATGACGAGAGAGCTGCTGAGAACTGACAGAGATATGATCTCTGAGACAGAACT 2349
 DB 499 ATTATGACGAGAGAGCTGCTGAGAACTGACAGAGATATGATCTCTGAGACAGAACT 440
 QY 2350 AAGATGCGAGATGAACAGACAGAGAGAG 2379
 DB 439 AAGATGCGAGATGAACAGACAGAGAGAG 410

RESULT 13 443 bp DNA linear GSS 30-JAN-2002
 BH618393/C Arabidopsis thaliana TDNA insertion lines Arabidopsis
 LOCUS SALK_039005 Arabidopsis thaliana genomic survey sequence.
 DEFINITION

ACCESSION BH618393.1 GI:18428488

VERSION GSS.
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana (thale cress); Embryophyta; Tracheophyta;
 Eukaryota; Viridiplantae; Streptophyta; Spermatophyta; Magnoliopsida; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 443)
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,
 Almonaco, J.M., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,
 Shinn, P., Zimmerman, J. and Ecker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome

JOURNAL Arabidopsis Genome
 COMMENT Unpublished (2001)
 Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@esalk.edu

FEATURES
 source 1..443
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_039005"

/clone_1b="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 5.6%; Score 184; DB 28; Length 443;
 Best Local Similarity 100.0%; Pred. No. 1.9e-80;
 Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1694 CTCCTGATCTTACGATCTTGAAGATGTTACATTTAGTGGCTGGCAGGCAACAG 1753
 DB 443 CTCCTGATCTTACGATCTTGAAGATGTTACATTTAGTGGCTGGCAGGCAACAG 384
 QY 1754 AGCTGCTGATATCTTGAAGATGTTACATTTAGTGGCTGGCAGGCAACAG 1813
 DB 383 AGCTGCTGATATCTTGAAGATGTTACATTTAGTGGCTGGCAGGCAACAG 324
 QY 1814 AGGCGCATCTGAGATGTTACATTTAGTGGCTGGCAGGCAACAG 1873
 DB 323 AGGCGCATCTGAGATGTTACATTTAGTGGCTGGCAGGCAACAG 264
 QY 1874 CCGA 1877
 DB 263 CCGA 260

RESULT 14 395 bp DNA linear GSS 02-JUL-2003
 BX547725 Arabidopsis thaliana T-DNA flanking sequence GK-547E01-020631,
 LOCUS genomic survey sequence.
 DEFINITION

ACCESSION BX547725.1 GI:32440534

VERSION GSS.
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana (thale cress); Embryophyta; Tracheophyta;
 Eukaryota; Viridiplantae; Streptophyta; Spermatophyta; Magnoliopsida; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 Strizhov, N., Li, Y., Rosso, M., Viehovec, P., Dekker, K., Saedler, H.
 and Weisshaar, B.

TITLE A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines

JOURNAL Unpublished
 COMMENT Unpublished

REFERENCE 2 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics

JOURNAL Unpublished
 COMMENT Unpublished

JOURNAL 3 (bases 1 to 395)
 Strizhov, N., Li, Y., Rosso, M. and Weisshaar, B.
 Direct Submission
 Submitted (01-JUL-2003) Weisshaar, B., Max-Planck-Institut fuer
 Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

TITLE This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone M0M1. The
 sequences are generated at the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
 http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES
 source 1..395
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"

/clone="GK-547E01-020631"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

ORIGIN

Query Match 5.2%; Score 170; DB 29; Length 395;
 Best Local Similarity 99.5%; Pred. No. 2.1e-73;
 Matches 220; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 163 AGCTTAATAAGCTTCCTCATTTGCTCTTCTTGGTCAGTTATTTCTTCTCCGAGATC 222
 |||||
 DB 50 AGCTTAATAAGCTTCCTCATTTGCTCTTCTTGGTCAGTTATTTCTTCTCCGAGATC 109
 |||||

QY 223 CTGACCTACTACTCTACTCTCCGGGCTTAACTTACGTTCCGCTGTTACTCTGT 282
 |||||
 DB 110 CTGACCTACTACTCTACTCTCCGGGCTTAACTTACGTTCCGCTGTTACTCTGT 169
 |||||

QY 283 AAGTTTTCGCTTGAAGCTCCGATCGCTCACCGCATCTTCTGTCGATTCTTC 342
 |||||
 DB 170 AAGTTTTCGCTTGAAGCTCCGATCGCTCACCGCATCTTCTGTCGATTCTTC 229
 |||||

QY 343 TTTTCTTCTGCTGAAAAATTTGCCCTAATGTTCTCGATTTC 383
 |||||
 DB 230 TTTTCTTCTGCTGAAAAATTTGCCCTAATGTTCTCGATTTC 270
 |||||

RESULT 15

BH169457 197 bp DNA linear GSS 03-OCT-2001
 LOCUS SALK_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis
 DEFINITION thaliana genomic clone SALK_001394, genomic survey sequence.

ACCESSION BH169457
 VERSION BH169457.1 GI:15904832

KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 197)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
 Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednisi,L.,
 Shin,P., Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome

JOURNAL COMMENT Unpublished (2001)
 CONTACT: Joseph R. Ecker
 The Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
 TDNA.

Class: TDNA tagged.

FEATURES Location/Qualifiers

source

1..197
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_001394"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines"

each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN

Query Match 4.7%; Score 153; DB 28; Length 197;
 Best Local Similarity 100.0%; Pred. No. 8.3e-65;
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 GTTATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGACGAGACTGCTTCTT 804
 |||||
 DB 1 GTTATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGACGAGACTGCTTCTT 60
 |||||

QY 805 CACAAGATGATGAGAGAGAGTGGAGTCATTTCCAGAAGAAACAAGAACAGAGAA 864
 |||||
 DB 61 CACAAGATGATGAGAGAGAGTGGAGTCATTTCCAGAAGAAACAAGAACAGAGAA 120
 |||||

QY 865 ACACTTCTGAAAAAACTTGGTTCTCAGAAAT 897
 |||||
 DB 121 ACACTTCTGAAAAAACTTGGTTCTCAGAAAT 153
 |||||

Search completed: April 9, 2004, 08:16:05
 Job time : 5513.97 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 20:17:10 ; Search time 161.43 Seconds
(without alignments)
11258.512 Million cell updates/sec

Title: US-10-030-829-1
Perfect score: 3275
Sequence: 1 gacaaacaaacaaacaaatataa.....caaatataggtgtacaaat 3275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	0.6	654	3 US-09-112-584-1	Sequence 1, Appl1
C 2	20	0.6	654	4 US-09-736-734-1	Sequence 1, Appl1
C 3	19	0.6	481	4 US-09-615-192A-109	Sequence 109, App
C 4	19	0.6	481	2 US-09-169-789-109	Sequence 109, App
C 5	19	0.6	495	2 US-08-975-316-27	Sequence 27, Appl
C 6	19	0.6	495	4 US-09-615-192A-27	Sequence 27, Appl
C 7	19	0.6	495	4 US-09-169-789-27	Sequence 27, Appl
C 8	19	0.6	1480	1 US-08-484-105-19	Sequence 19, Appl
C 9	19	0.6	1480	1 US-08-484-106-19	Sequence 19, Appl
C 10	19	0.6	1857	3 US-09-377-557-9	Sequence 9, Appl1
C 11	19	0.6	2000	1 US-09-041-075A-4	Sequence 4, Appl1
C 12	19	0.6	2000	1 US-09-041-075A-6	Sequence 6, Appl1
C 13	19	0.6	8257	4 US-09-595-684B-30	Sequence 30, Appl1
C 14	19	0.6	8503	4 US-09-620-312D-110	Sequence 130, Appl
C 15	19	0.6	9531	4 US-09-163-748C-3	Sequence 3, Appl1
C 16	19	0.6	10254	4 US-08-961-527-29	Sequence 29, Appl
C 17	19	0.6	11461	4 US-08-669-161A-29	Sequence 29, Appl
C 18	19	0.6	22846	2 US-08-469-461-3	Sequence 3, Appl1
C 19	19	0.6	22846	3 US-07-890-609-3	Sequence 3, Appl1
C 20	19	0.6	392000	4 US-10-027-983-11	Sequence 11, Appl1
C 21	19	0.5	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
C 22	18	0.5	302	4 US-09-313-294A-488	Sequence 488, App
C 23	18	0.5	346	4 US-09-621-976-17456	Sequence 17456, A
C 24	18	0.5	441	4 US-09-621-976-10425	Sequence 10425, A
C 25	18	0.5	462	4 US-09-489-039A-5315	Sequence 5315, App
C 26	18	0.5	608	3 US-09-385-982-236	Sequence 236, App
C 27	18	0.5	694	1 US-08-250-314-1	Sequence 1, Appl1

C 28	18	0.5	694	1 US-08-708-107-1	Sequence 1, Appl1
C 29	18	0.5	947	4 US-08-958-009D-11	Sequence 11, Appl
C 30	18	0.5	947	4 US-08-958-009D-12	Sequence 12, Appl
C 31	18	0.5	948	4 US-08-958-009D-6	Sequence 6, Appl1
C 32	18	0.5	948	4 US-08-958-009D-7	Sequence 7, Appl1
C 33	18	0.5	948	4 US-08-958-009D-10	Sequence 10, Appl
C 34	18	0.5	948	4 US-08-958-009D-13	Sequence 13, Appl
C 35	18	0.5	948	4 US-08-958-009D-14	Sequence 14, Appl
C 36	18	0.5	950	4 US-08-958-009D-8	Sequence 8, Appl1
C 37	18	0.5	950	4 US-08-958-009D-15	Sequence 15, Appl
C 38	18	0.5	960	4 US-09-134-001C-2618	Sequence 2618, Ap
C 39	18	0.5	1254	4 US-09-322-478-25	Sequence 25, Appl
C 40	18	0.5	1381	2 US-08-454-557C-49	Sequence 49, Appl
C 41	18	0.5	1381	2 US-08-340-426D-49	Sequence 49, Appl
C 42	18	0.5	1381	2 US-08-450-673C-49	Sequence 49, Appl
C 43	18	0.5	1381	5 PCT-US95-17111A-49	Sequence 49, Appl
C 44	18	0.5	1418	5 PCT-US95-17111A-120	Sequence 120, App
C 45	18	0.5	1442	2 US-08-454-557C-120	Sequence 120, App

ALIGNMENTS

```

RESULT 1
US-09-112-584-1/c
; Sequence 1, Application US/09112584
; Patent No. 6228638
; GENERAL INFORMATION:
; APPLICANT: Romeo, Tony
; TITLE OF INVENTION: THE ESCHERICHIA COLI CSRB GENE, RNA ENCODED THEREBY,
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 31608200200
; CURRENT APPLICATION NUMBER: US/09/112,584
; CURRENT FILING DATE: 1998-07-09
; EARLIER APPLICATION NUMBER: US 60/052,372
; EARLIER FILING DATE: 1997-07-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-112-584-1

Query Match      0.6%; Score 20; DB 3; Length 654;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3164 AAAATTTCATGGAAGTTA 3183
Db      153 AAAATTTCATGGAAGTTA 134

RESULT 2.
US-09-736-734-1/c
; Sequence 1, Application US/09736734
; Patent No. 6537815
; GENERAL INFORMATION:
; APPLICANT: Romeo, Tony
; TITLE OF INVENTION: ESCHERICHIA COLI CSRB GENE, RNA ENCODED
; TITLE OF INVENTION: THEREBY, AND METHODS OF USE THEREOF
; FILE REFERENCE: 316082002001
; CURRENT APPLICATION NUMBER: US/09/736,734
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: 09/112,584
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/052,372
; PRIOR FILING DATE: 1997-07-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 654

```

TYPE: DNA
ORGANISM: Escherichia coli
US-09-736-734-1

Query Match 0.6%; Score 20; DB 4; Length 654;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTCATGAAAGTTA 3183
DB 153 AAAATTCATGAAAGTTA 134

RESULT 3
US-09-615-192A-109/c

Sequence 109, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003C4U
CURRENT APPLICATION NUMBER: US/09/615,192A
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-109

Query Match 0.6%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787
DB 470 TGGTCAAGTTGGCAGG 452

RESULT 4
US-09-169-789-109/c

Sequence 109, Application US/09169789
Patent No. 6653528
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003C2
CURRENT APPLICATION NUMBER: US/09/169,789
PRIOR FILING DATE: 1998-10-09
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-109

Query Match 0.6%; Score 19; DB 4; Length 481;

Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787
DB 470 TGGTCAAGTTGGCAGG 452

RESULT 5
US-08-975-316-27/c

Sequence 27, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N.; HAVUKKALA, Ilkka
APPLICANT: GRIERSON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESSES:
ADDRESSER: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-27

Query Match 0.6%; Score 19; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787
DB 484 TGGTCAAGTTGGCAGG 466

RESULT 6
US-09-615-192A-27/c

Sequence 27, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003C4U
CURRENT APPLICATION NUMBER: US/09/615,192A

;; CURRENT FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 08/975,316
;; PRIOR FILING DATE: 1997-11-21
;; PRIOR APPLICATION NUMBER: US 08/713,000
;; PRIOR FILING DATE: 1996-09-11
;; PRIOR APPLICATION NUMBER: US 09/169,789
;; PRIOR FILING DATE: 1998-10-09
;; NUMBER OF SEQ ID NOS: 405
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 27
;; LENGTH: 495
;; TYPE: DNA
;; ORGANISM: Eucalyptus grandis
US-09-615-192A-27

Query Match 0.6%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTGGCAGG 787
DB 484 TGGTTCAGGTTGGCAGG 466

RESULT 7
US-09-169-789-27/c
; Sequence 27, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Blockberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c2
; CURRENT APPLICATION NUMBER: US/09/169,789
; EARLIER FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 08/975,316
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: US 08/713,000
; EARLIER FILING DATE: 1996-09-11
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 27
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-169-789-27

Query Match 0.6%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTGGCAGG 787
DB 484 TGGTTCAGGTTGGCAGG 466

RESULT 8
US-08-484-105-19
; Sequence 19, Application US/08484105
; Patent No. 5589341
; GENERAL INFORMATION:
; APPLICANT: STILLMAN, Bruce
; APPLICANT: BELL, Stephen P
; APPLICANT: KOBAYASHI, Ryuji
; APPLICANT: RINE, Jasper
; APPLICANT: MCNALLY, Margit
; APPLICANT: MCNALLY, Francis J
; APPLICANT: LAURENSEN, Patricia
; APPLICANT: HERSKOWITZ, Ira
; APPLICANT: LI, Joachim J
; APPLICANT: GAVIN, Kimberly
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
;; STREET: 4 Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-4187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/484,105
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Osman Ph.D., Richard Aron
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 494-8700
;; TELEFAX: (415) 494-8771
;; TELEX: 910 277299
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1480 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 277..1365
US-08-484-105-19

Query Match 0.6%; Score 19; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 TAAGTGAATTCCTCTGT 1673
DB 238 TAAGTGAATTCCTCTGT 256

RESULT 9
US-08-484-106-19
; Sequence 19, Application US/08484106
; Patent No. 5614618
; GENERAL INFORMATION:
; APPLICANT: STILLMAN, Bruce
; APPLICANT: BELL, Stephen P
; APPLICANT: KOBAYASHI, Ryuji
; APPLICANT: RINE, Jasper
; APPLICANT: MCNALLY, Margit
; APPLICANT: MCNALLY, Francis J
; APPLICANT: LAURENSEN, Patricia
; APPLICANT: HERSKOWITZ, Ira
; APPLICANT: LI, Joachim J
; APPLICANT: GAVIN, Kimberly
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATE:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/PAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 277..1365
US-08-484-106-19

```

```

Query Match          0.6%; Score 19; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1655 TAAGTGAATTCCTCTGT 1673
DB      238 TAAGTGAATTCCTCTGT 256

```

```

RESULT 10
US-09-377-557-9/c
Sequence 9, Application US/09377557
Patent No. 6297055
GENERAL INFORMATION:
APPLICANT: Falco, S. Carl
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Emil M. Jr.
TITLE OF INVENTION: Amino Acid Decarboxylases
FILE REFERENCE: BB-1237
CURRENT APPLICATION NUMBER: US/09/377,557
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 60/099,493
EARLIER FILING DATE: September 8, 1998
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 1857
TYPE: DNA
ORGANISM: Oryza sativa
US-09-377-557-9

```

```

Query Match          0.6%; Score 19; DB 3; Length 1857;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3152 TTTTGTGTAATTT 3170
DB      1844 TTTTGTGTAATTT 1826

```

```

RESULT 11
US-09-041-075A-4
Sequence 4, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A
APPLICANT: Radding, Jeffrey A

```

```

TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 2000
TYPE: DNA
ORGANISM: Candida krusei
FEATURE:
NAME/KEY: CDS
LOCATION: (300)..(1739)
US-09-041-075A-4

```

```

Query Match          0.6%; Score 19; DB 1; Length 2000;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1218 GATGATTCGATGACGACC 1236
DB      1560 GATGATTCGATGACGACC 1578

```

```

RESULT 12
US-09-041-075A-6
Sequence 6, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A
APPLICANT: Radding, Jeffrey A
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 2000
TYPE: RNA
ORGANISM: Candida krusei
US-09-041-075A-6

```

```

Query Match          0.6%; Score 19; DB 1; Length 2000;
Best Local Similarity 73.7%; Pred. No. 25;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1218 GATGATTCGATGACGACC 1236
DB      1560 GAUGAUGAUGACGACC 1578

```

```

RESULT 13
US-09-595-684B-30
Sequence 30, Application US/09595684B
Patent No. 6544766
GENERAL INFORMATION:
APPLICANT: Bertrand, Christophe
APPLICANT: Ohashi, Cara
APPLICANT: Sakowicz, Roman
APPLICANT: Valsberg, Eugene
APPLICANT: Wood, Kenneth
APPLICANT: Yu, Ming
TITLE OF INVENTION: Human kinesins and methods of producing
TITLE OF INVENTION: Human kinesins and purifying human kinesins
FILE REFERENCE: cytop036
CURRENT APPLICATION NUMBER: US/09/595,684B

```

;; CURRENT FILING DATE: 2002-06-24
;; PRIOR APPLICATION NUMBER: 09/295,612
;; PRIOR FILING DATE: 2000-04-20
;; NUMBER OF SEQ ID NOS: 105
;; SOFTWARE: FastSeq for windows Version 4.0
;; SEQ ID NO 30
;; LENGTH: 8257
;; TYPE: DNA
;; ORGANISM: Human
US-09-595-684B-30

Query Match 0.6%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2873 TTGATCTGGAGAAAGATT 2891
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 14
US-09-620-312D-130
;; Sequence 130, Application US/09620312D
;; Patent No. 6569662
;; GENERAL INFORMATION:
;; APPLICANT: Tang, Y. Tom
;; APPLICANT: Liu, Chenghua
;; APPLICANT: Asundi, Vinod
;; APPLICANT: Zhang, Jie
;; APPLICANT: Ren, Feiyan
;; APPLICANT: Chen, Rui-hong
;; APPLICANT: Zhao, Qing A.
;; APPLICANT: Wehrman, Tom
;; APPLICANT: Xue, Aidong J.
;; APPLICANT: Yang, Yonghong
;; APPLICANT: Wang, Jian-Rui
;; APPLICANT: Zhou, Ping
;; APPLICANT: Ma, Yundong
;; APPLICANT: Wang, Dunrui
;; APPLICANT: Wang, Zhiwei
;; APPLICANT: John Tillinghast
;; APPLICANT: Drmanac, Radoje T.
;; TITLE OF INVENTION: No. 6569662zel Nucleic Acids and
;; FILE REFERENCE: 784CIP28
;; CURRENT APPLICATION NUMBER: US/09/620,312D
;; CURRENT FILING DATE: 2000-07-19
;; PRIOR APPLICATION NUMBER: 09/552,317
;; PRIOR FILING DATE: 2000-04-25
;; PRIOR APPLICATION NUMBER: 09/488,725
;; PRIOR FILING DATE: 2000-01-21
;; NUMBER OF SEQ ID NOS: 1105
;; SOFTWARE: pc_fl_genes Version 1.0
;; SEQ ID NO 130
;; LENGTH: 8503
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (91)..(8082)
US-09-620-312D-130

Query Match 0.6%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2873 TTGATCTGGAGAAAGATT 2891
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 15
US-09-163-748C-3/c

;; Sequence 3, Application US/09163748C
;; Patent No. 6509172
;; GENERAL INFORMATION:
;; APPLICANT: DeBacker, Oliver
;; APPLICANT: Van den Eynde, Benoit
;; APPLICANT: Boon-Falleur, Thierry
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode Member Of The Gage
;; FILE REFERENCE: LUD 5558
;; CURRENT APPLICATION NUMBER: US/09/163,748C
;; CURRENT FILING DATE: 1998-09-30
;; NUMBER OF SEQ ID NOS: 27
;; SEQ ID NO 3
;; LENGTH: 9531
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: 92,232,1041,7412,9038-9040
;; OTHER INFORMATION: Identity of several nucleotides not known
US-09-163-748C-3

Query Match 0.6%; Score 19; DB 4; Length 9531;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACAAAATGAAACAAA 109
DB 7045 TACAAAATGAAACAAA 7027

Search completed: April 9, 2004, 08:20:42
Job time : 165.43 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 9, 2004, 05:52:11 ; Search time 923.457 Seconds
(without alignments)
13304.392 Million cell updates/sec

Title: US-10-030-829-1
- Perfect score: 3275
Sequence: 1 gcaacaacaacaataa.....caaatataggtacaaat 3275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2475585 seqs, 1875730760 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	0.7	6136	US-10-257-166-123	Sequence 123, App
2	22	0.7	9888	US-10-311-455-1214	Sequence 1214, Ap
3	21	0.6	2164	US-10-027-633-250839	Sequence 250839,
4	21	0.6	6048	US-10-311-455-482	Sequence 482, App
5	21	0.6	6072	US-10-311-455-4	Sequence 4, Appli
6	21	0.6	8245	US-10-221-714A-170	Sequence 170, App
7	20	0.6	476	US-10-001-843-87	Sequence 87, Appl
8	20	0.6	654	US-09-736-734-1	Sequence 1, Appli
9	20	0.6	654	US-10-396-911-1	Sequence 133, App
10	20	0.6	1023	US-10-074-475-133	Sequence 91177, A
11	20	0.6	1377	US-10-424-599-91177	Sequence 10949, A
12	20	0.6	1461	US-10-282-122A-10949	Sequence 91176, A
13	20	0.6	1589	US-10-424-599-91176	Sequence 1756, Ap
14	20	0.6	2418	US-10-108-260A-1756	Sequence 1, Appli
15	20	0.6	2985	US-09-947-953-1	Sequence 1, Appli

16	20	0.6	5278	12	US-10-221-714A-98	Sequence 98, Appl
17	20	0.6	5278	14	US-10-311-455-796	Sequence 796, App
18	20	0.6	5376	14	US-10-311-455-2123	Sequence 2123, Ap
19	20	0.6	5453	12	US-10-221-613-176	Sequence 376, App
20	20	0.6	6077	14	US-10-311-455-1828	Sequence 1828, Ap
21	20	0.6	6212	14	US-10-240-453-7	Sequence 7, Appli
22	20	0.6	6412	12	US-10-221-714A-353	Sequence 353, App
23	20	0.6	6412	14	US-10-311-455-1883	Sequence 1883, Ap
24	20	0.6	7624	14	US-10-311-455-2085	Sequence 2085, Ap
25	20	0.6	10696	14	US-10-311-455-291	Sequence 291, App
26	20	0.6	10696	14	US-10-240-452-39	Sequence 39, Appl
27	20	0.6	11422	14	US-10-311-455-191	Sequence 191, App
28	20	0.6	11422	16	US-10-257-166-17	Sequence 17, Appl
29	20	0.6	13732	20	US-10-311-455-1793	Sequence 1793, Ap
30	20	0.6	14708	12	US-10-221-714A-499	Sequence 499, App
31	20	0.6	14708	14	US-10-239-676-221	Sequence 221, App
32	20	0.6	14708	14	US-10-311-455-2217	Sequence 2217, Ap
33	20	0.6	14708	14	US-10-240-453-323	Sequence 323, App
34	20	0.6	16724	14	US-10-311-455-1064	Sequence 1064, Ap
35	20	0.6	16724	14	US-10-240-485-90	Sequence 90, Appl
36	20	0.6	17580	14	US-10-240-452-12	Sequence 12, Appl
37	20	0.6	17993	9	US-09-768-781-5	Sequence 5, Appli
38	20	0.6	34769	12	US-10-221-714A-502	Sequence 502, App
39	20	0.6	73334	14	US-10-311-455-2098	Sequence 2098, Ap
40	20	0.6	302603	12	US-10-271-416-8	Sequence 8, Appli
41	20	0.6	3673778	14	US-10-312-841-1	Sequence 1, Appli
42	20	0.6	3673778	14	US-10-312-841-2	Sequence 2, Appli
43	19	0.6	172	12	US-10-085-783A-19246	Sequence 19246, A
44	19	0.6	172	15	US-10-242-535A-19246	Sequence 19246, A
45	19	0.6	385	14	US-10-066-543-1881	Sequence 1881, Ap

ALIGNMENTS

RESULT 1
US-10-257-166-123
; Sequence 123, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBERCK, Christian
; APPLICANT: BERLIN, Kutt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 123
; LENGTH: 6136
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-123

Query Match 0.7%; Score 22; DB 16; Length 6136;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3145 TATTTGTTTTTTTGTGTA 3166
|||||
Db 2122 TATTTGTTTTTTTGTGTA 2143

RESULT 2

US-10-311-455-1214
; Sequence 1214; Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1214
; LENGTH: 9888
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1214

Query Match 0.7%; Score 22; DB 14; Length 9888;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3142 TAGTATTTTGTGTTTTTGT 3163
DB 8219 TAGTATTTTGTGTTTTTGT 8240

RESULT 3
US-10-027-632-250839
; Sequence 250839; Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: PASTSEQ for Windows Version 4.0
; SEQ ID NO 250839
; LENGTH: 2164
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250839

Query Match 0.6%; Score 21; DB 15; Length 2164;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3154 TTTTGTGTAATTTTCATA 3174
DB 842 TTTTGTGTAATTTTCATA 862

RESULT 4
US-10-311-455-482
; Sequence 482; Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 482
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-482

Query Match 0.6%; Score 21; DB 14; Length 6048;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 526 TTCGTTTTTGTGTTTTGTT 546
DB 5395 TTCGTTTTTGTGTTTTGTT 5415

RESULT 5
US-10-311-455-4
; Sequence 4; Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 4
; LENGTH: 6072
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-4

Query Match 0.6%; Score 21; DB 14; Length 6072;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTTCATATGAAAGTTA 3183
DB 153 AAAATTTCATATGAAAGTTA 134

RESULT 10
US-10-074-475-133
; Sequence 133, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Reipon, Hervé
; APPLICANT: Kaita, Kalpana
; APPLICANT: Cafierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0313
; CURRENT APPLICATION NUMBER: US/10/074,475
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 60/268,292
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 295
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 133
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-074-475-133

Query Match 0.6%; Score 20; DB 14; Length 1023;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3136 TTTCCTTAGATTTGTTT 3155
DB 832 TTTCCTTAGATTTGTTT 851

RESULT 11
US-10-424-599-91177
; Sequence 91177, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91177
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177

Query Match 0.6%; Score 20; DB 12; Length 1377;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCCAGCTTG 1387

DB 359 CAGTGCATTGTCCAGCTTG 378

RESULT 12
US-10-282-122A-10949/C
; Sequence 10949, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10949
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-10949

Query Match 0.6%; Score 20; DB 12; Length 1461;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2047 ATATATGCTTTAGTTTGT 2066
DB 1073 ATATATGCTTTAGTTTGT 1054

RESULT 13
US-10-424-599-91176
; Sequence 91176, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 91176
 LENGTH: 1589
 TYPE: DNA
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
 US-10-424-599-91176

Query Match 0.6%; Score 20; DB 12; Length 1589;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCAGCTTG 1387
 DB 359 CAGTGCATTGTCAGCTTG 378

RESULT 14
 US-10-108-260A-1756
 Sequence 1756, Application US/10108260A
 Publication No. US20040005560A1
 GENERAL INFORMATION:
 APPLICANT: HELIX RESEARCH INSTITUTE
 TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
 FILE REFERENCE: H1-A0106
 CURRENT APPLICATION NUMBER: US/10/108,260A
 CURRENT FILING DATE: 2002-03-27
 NUMBER OF SEQ ID NOS: 5458
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1756
 LENGTH: 2418
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-108-260A-1756

Query Match 0.6%; Score 20; DB 15; Length 2418;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3165 AAATTCATATGAAAGTAG 3184
 DB 1988 AAATTCATATGAAAGTAG 2007

RESULT 15
 US-09-947-953-1/C
 Sequence 1, Application US/09947953
 Patent No. US20020155101A1
 GENERAL INFORMATION:
 APPLICANT: DONAHUE, J. KEVIN
 APPLICANT: MARBAN, EDUARDO
 TITLE OF INVENTION: CARDIAC ARRHYTHMIA TREATMENT METHODS
 FILE REFERENCE: 71699/56415
 CURRENT APPLICATION NUMBER: US/09/947,953
 CURRENT FILING DATE: 2001-09-06
 PRIOR APPLICATION NUMBER: 60/230,311
 PRIOR FILING DATE: 2001-09-06
 PRIOR APPLICATION NUMBER: 60/295,889
 PRIOR FILING DATE: 2001-06-05
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 2985
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-947-953-1

Query Match 0.6%; Score 20; DB 9; Length 2985;

Best Local Similarity 100.0%; Pred. No. 38;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CGAAAAAGGAAAAAAGG 90
 DB 25 CGAAAAAGGAAAAAAGG 6

Search completed: April 9, 2004, 11:58:46
 Job time : 929.457 secs

This Page Blank (uspto)

JOURNAL

Patent: WO 0105951-A 2 25-JAN-2001;
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)

FEATURES

source

Location/Qualifiers

1. .1878
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 1. .1878
 /note="unnamed protein product"

CDS

/codon_start=1
 /protein_id="CAC32419.1"
 /db_xref="GI:13158381"
 /db_xref="REMBL:CA032419"
 /translation="MSSRAGPMSKKNVQGYREYVQGLAGTRILASSODDGGEM
 EVISKKNKRGNTSGKTWVSONSPRAMGQOQGRSNVGRGNVSGRNGRG
 IONISGRGRLARKYDNNFVAPPVSRPPLGGMMWQAGSAQHTVQEPVEDD
 VDNABENDSDALDSDDDLADSDVDSOKSHGRKQNKFFKFGSLDLSIEQ
 INEPRQWHPACONPGALDWNLHPLAHATKARVYLKRELAELEKDLQMRG
 ASVLPCEIYGQWKGEDEKDYETWPEPVYIEMRLKNDNDKWLGMGNOLLEYF
 DKYEALRARSYGPQGHKRWVLMFSSATGIEMRLKRELAELEKDLQMRG
 FSGVROLYGFLLTKDLDIFNQSOKTRKLFELSYQPMYKELROI SEMOOLNY
 FKMLSKONKHAUVLESLIMSEKLRTEENRIVROKMQHEONREMAHREF
 MDSIKOHERDRAKEENFEMLOQERAKVQOQONINPSDDCKRAEYVSTIER
 QEKMEFVERHEMLIKDEKQKMDMKRHEEIPDLKEFDEALQLMYKGLHNE
 D"

ORIGIN

Query Match 100.0%; Score 1878; DB 6; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTCTAGAGGCTGGTCCATGCTTAAGAAAAAGAACTTCAGGCTGTTATAGCCCT 60
 1 ATGAGTTCTAGAGGCTGGTCCATGCTTAAGAAAAAGAACTTCAGGCTGTTATAGCCCT 60
 61 GAGGTGAACAGTGGTTCAGGTTTGGCAGGAGCAGACCTGGCTTCTTCAAGATGAT 120
 61 GAGGTGAACAGTGGTTCAGGTTTGGCAGGAGCAGACCTGGCTTCTTCAAGATGAT 120
 121 GAGAGAGTGGAGGCTCACTTCCAAAGAAACAAGAACAAACAGAAACACTTCTGGA 180
 121 GAGAGAGTGGAGGCTCACTTCCAAAGAAACAAGAACAAACAGAAACACTTCTGGA 180
 122 GAGAGAGTGGAGGCTCACTTCCAAAGAAACAAGAACAAACAGAAACACTTCTGGA 180
 181 AAAAATTGGGTTCTCAGATTGCAATCCTCTAGACTTGGGCTGTCAGCAGAAAGG 240
 181 AAAAATTGGGTTCTCAGATTGCAATCCTCTAGACTTGGGCTGTCAGCAGAAAGG 240
 181 AAAAATTGGGTTCTCAGATTGCAATCCTCTAGACTTGGGCTGTCAGCAGAAAGG 240
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATCCGGGAGAGGTAAACGCAATGAT 300
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATCCGGGAGAGGTAAACGCAATGAT 300
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATCCGGGAGAGGTAAACGCAATGAT 300
 301 CGGGGCACTTCAAGCTACATATCTGGTCCGGGAGCAGACGCTTGGAGCAAAAGTATGATAC 360
 301 CGGGGCACTTCAAGCTACATATCTGGTCCGGGAGCAGACGCTTGGAGCAAAAGTATGATAC 360
 301 CGGGGCACTTCAAGCTACATATCTGGTCCGGGAGCAGACGCTTGGAGCAAAAGTATGATAC 360
 301 CGGGGCACTTCAAGCTACATATCTGGTCCGGGAGCAGACGCTTGGAGCAAAAGTATGATAC 360
 361 AACTTGTGGCAACCCCACTGTATCTCCGCTCTTTGGAAGAGAGATGGAATGGCAG 420
 361 AACTTGTGGCAACCCCACTGTATCTCCGCTCTTTGGAAGAGAGATGGAATGGCAG 420
 361 AACTTGTGGCAACCCCACTGTATCTCCGCTCTTTGGAAGAGAGATGGAATGGCAG 420
 361 AACTTGTGGCAACCCCACTGTATCTCCGCTCTTTGGAAGAGAGATGGAATGGCAG 420
 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTTCTGACGTGGAGATGAT 480
 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTTCTGACGTGGAGATGAT 480
 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTTCTGACGTGGAGATGAT 480
 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTTCTGACGTGGAGATGAT 480
 481 GTGATTAATGCTTCTGAGAAAGAGATGATCCGATGCTTTGATGATGATGATGACGAC 540
 481 GTGATTAATGCTTCTGAGAAAGAGATGATCCGATGCTTTGATGATGATGATGATGACGAC 540
 481 GTGATTAATGCTTCTGAGAAAGAGATGATCCGATGCTTTGATGATGATGATGATGACGAC 540
 481 GTGATTAATGCTTCTGAGAAAGAGATGATCCGATGCTTTGATGATGATGATGATGACGAC 540
 541 CTTGCAAGATGATTAATGATCTCGAGATGATGATCAAAAGACCATGATGATCGAAAGCAG 600
 541 CTTGCAAGATGATTAATGATCTCGAGATGATGATCAAAAGACCATGATGATGATCGAAAGCAG 600
 541 CTTGCAAGATGATTAATGATCTCGAGATGATGATCAAAAGACCATGATGATGATCGAAAGCAG 600
 541 CTTGCAAGATGATTAATGATCTCGAGATGATGATCAAAAGACCATGATGATGATCGAAAGCAG 600
 601 AATAAGTGGTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGATGATGAT 660

601 AATAAGTGGTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGATGATGATGAT 660
 601 AATAAGTGGTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGATGATGATGAT 660
 661 GAACACAGAGGAGTGGCATTTGTCAGCTTGTGACAAACGACCTGGTGCATCATGATGG 720
 661 GAACACAGAGGAGTGGCATTTGTCAGCTTGTGACAAACGACCTGGTGCATCATGATGG 720
 661 GAACACAGAGGAGTGGCATTTGTCAGCTTGTGACAAACGACCTGGTGCATCATGATGG 720
 661 GAACACAGAGGAGTGGCATTTGTCAGCTTGTGACAAACGACCTGGTGCATCATGATGG 720
 721 TATTAACCTGCACCTCTTACTAGCTCATAGCGAGGACAAAGAGGCTAGGCGATTAAGCTC 780
 721 TATTAACCTGCACCTCTTACTAGCTCATAGCGAGGACAAAGAGGCTAGGCGATTAAGCTC 780
 721 TATTAACCTGCACCTCTTACTAGCTCATAGCGAGGACAAAGAGGCTAGGCGATTAAGCTC 780
 721 TATTAACCTGCACCTCTTACTAGCTCATAGCGAGGACAAAGAGGCTAGGCGATTAAGCTC 780
 721 TATTAACCTGCACCTCTTACTAGCTCATAGCGAGGACAAAGAGGCTAGGCGATTAAGCTC 780
 781 CATTAAGAAATTGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGGCGATCTGTCACT 840
 781 CATTAAGAAATTGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGGCGATCTGTCACT 840
 781 CATTAAGAAATTGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGGCGATCTGTCACT 840
 781 CATTAAGAAATTGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGGCGATCTGTCACT 840
 841 CTTTGTGTGAGATTTTATGCGGAGTGAAGAGGTTTGGGTGAGATGAAGAAAGATTAAGAA 900
 841 CTTTGTGTGAGATTTTATGCGGAGTGAAGAGGTTTGGGTGAGATGAAGAAAGATTAAGAA 900
 841 CTTTGTGTGAGATTTTATGCGGAGTGAAGAGGTTTGGGTGAGATGAAGAAAGATTAAGAA 900
 841 CTTTGTGTGAGATTTTATGCGGAGTGAAGAGGTTTGGGTGAGATGAAGAAAGATTAAGAA 900
 841 CTTTGTGTGAGATTTTATGCGGAGTGAAGAGGTTTGGGTGAGATGAAGAAAGATTAAGAA 900
 901 ATTGTCTGGGCTGCATAGGTCATCATGAAATTAATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGGGCTGCATAGGTCATCATGAAATTAATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGGGCTGCATAGGTCATCATGAAATTAATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGGGCTGCATAGGTCATCATGAAATTAATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGGGCTGCATAGGTCATCATGAAATTAATGATGATGATGATGATGATGATGAT 960
 961 AAGTGGCTGGCAATGGGCAACAAAGAGCTGCTGGAATACTTGGAGCAAGATGATGATGAT 1020
 961 AAGTGGCTGGCAATGGGCAACAAAGAGCTGCTGGAATACTTGGAGCAAGATGATGATGAT 1020
 961 AAGTGGCTGGCAATGGGCAACAAAGAGCTGCTGGAATACTTGGAGCAAGATGATGATGAT 1020
 961 AAGTGGCTGGCAATGGGCAACAAAGAGCTGCTGGAATACTTGGAGCAAGATGATGATGAT 1020
 961 AAGTGGCTGGCAATGGGCAACAAAGAGCTGCTGGAATACTTGGAGCAAGATGATGATGAT 1020
 1021 AGAGCAGGCCATTCCTATGCTCCACAGAGGCTGCTGAGATGATGATGATGATGATGATGAT 1080
 1021 AGAGCAGGCCATTCCTATGCTCCACAGAGGCTGCTGAGATGATGATGATGATGATGATGAT 1080
 1021 AGAGCAGGCCATTCCTATGCTCCACAGAGGCTGCTGAGATGATGATGATGATGATGATGAT 1080
 1021 AGAGCAGGCCATTCCTATGCTCCACAGAGGCTGCTGAGATGATGATGATGATGATGATGAT 1080
 1081 AGCAGTGCATGCTGCTATTTGAGGCGGCAACGCTTCCACCGGAGTATGATGATGATGATG 1140
 1081 AGCAGTGCATGCTGCTATTTGAGGCGGCAACGCTTCCACCGGAGTATGATGATGATGATG 1140
 1081 AGCAGTGCATGCTGCTATTTGAGGCGGCAACGCTTCCACCGGAGTATGATGATGATGATG 1140
 1081 AGCAGTGCATGCTGCTATTTGAGGCGGCAACGCTTCCACCGGAGTATGATGATGATGATG 1140
 1141 TTAGATTAATTTGCTGGGCTGAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200
 1141 TTAGATTAATTTGCTGGGCTGAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200
 1141 TTAGATTAATTTGCTGGGCTGAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200
 1141 TTAGATTAATTTGCTGGGCTGAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200
 1141 TTAGATTAATTTGCTGGGCTGAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200
 1201 TATGCTTCTTGCACACAGAGAGATCTGAGCAATTAATCAACTCTCAAGGCAAA 1260
 1201 TATGCTTCTTGCACACAGAGAGATCTGAGCAATTAATCAACTCTCAAGGCAAA 1260
 1201 TATGCTTCTTGCACACAGAGAGATCTGAGCAATTAATCAACTCTCAAGGCAAA 1260
 1201 TATGCTTCTTGCACACAGAGAGATCTGAGCAATTAATCAACTCTCAAGGCAAA 1260
 1261 ACAAGGCTGAATTTGAGTGAATTAATCAATCAAGAGATGTTTGAAGAGAGCTGAGCAG 1320
 1261 ACAAGGCTGAATTTGAGTGAATTAATCAATCAAGAGATGTTTGAAGAGAGCTGAGCAG 1320
 1261 ACAAGGCTGAATTTGAGTGAATTAATCAATCAAGAGATGTTTGAAGAGAGCTGAGCAG 1320
 1261 ACAAGGCTGAATTTGAGTGAATTAATCAATCAAGAGATGTTTGAAGAGAGCTGAGCAG 1320
 1261 ACAAGGCTGAATTTGAGTGAATTAATCAATCAAGAGATGTTTGAAGAGAGCTGAGCAG 1320
 1321 ATCTCTGAGGACATGACAGCTGAACTTAAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTCTGAGGACATGACAGCTGAACTTAAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTCTGAGGACATGACAGCTGAACTTAAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTCTGAGGACATGACAGCTGAACTTAAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTCTGAGGACATGACAGCTGAACTTAAAGAACAGCTTCAAAACAGAACAG 1380
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGGAGAGCTGCTGAACTGCA 1440
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGGAGAGCTGCTGAACTGCA 1440
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGGAGAGCTGCTGAACTGCA 1440
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGGAGAGCTGCTGAACTGCA 1440
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATGAGGAGAGCTGCTGAACTGCA 1440
 1441 GAGGATTAATGATGATGAG 1500
 1441 GAGGATTAATGATGATGAG 1500
 1441 GAGGATTAATGATGATGAG 1500
 1441 GAGGATTAATGATGATGAG 1500
 1441 GAGGATTAATGATGATGAG 1500
 1501 ATGATGACACACAGCAGGTTTTCATGATTAATCAATCAAGATCAATGAAGAGAGAGC 1560
 1501 ATGATGACACACAGCAGGTTTTCATGATTAATCAATCAAGATCAATGAAGAGAGAGC 1560
 1501 ATGATGACACACAGCAGGTTTTCATGATTAATCAATCAAGATCAATGAAGAGAGAGC 1560
 1501 ATGATGACACACAGCAGGTTTTCATGATTAATCAATCAAGATCAATGAAGAGAGAGC 1560
 1501 ATGATGACACACAGCAGGTTTTCATGATTAATCAATCAAGATCAATGAAGAGAGAGC 1560
 1561 GCAAG 1620
 1561 GCAAG 1620
 1561 GCAAG 1620
 1561 GCAAG 1620
 1561 GCAAG 1620
 1621 CAGAGCAGAGACATTAATCCCTTACGAAATGACGATGCTGGAAGAGAGAGAGAGAG 1680
 1621 CAGAGCAGAGACATTAATCCCTTACGAAATGACGATGCTGGAAGAGAGAGAGAGAG 1680
 1621 CAGAGCAGAGACATTAATCCCTTACGAAATGACGATGCTGGAAGAGAGAGAGAGAG 1680
 1621 CAGAGCAGAGACATTAATCCCTTACGAAATGACGATGCTGGAAGAGAGAGAGAGAG 1680
 1621 CAGAGCAGAGACATTAATCCCTTACGAAATGACGATGCTGGAAGAGAGAGAGAGAG 1680
 1681 TCAAGCTTCAAGGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 1681 TCAAGCTTCAAGGTTTCAAG 1740

QY 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGATAGCTTGTGATGAGAGATTAAT 660
 Db 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGATAGCTTGTGATGAGAGATTAAT 660
 QY 661 GAACCAAGAGGAGGAGATTTGTCAGAAAGAGCTGTGTCCATGATGAG 720
 Db 661 GAACCAAGAGGAGGAGATTTGTCAGAAAGAGCTGTGTCCATGATGAG 720
 QY 661 GAACCAAGAGGAGGAGATTTGTCAGAAAGAGCTGTGTCCATGATGAG 720
 Db 661 GAACCAAGAGGAGGAGATTTGTCAGAAAGAGCTGTGTCCATGATGAG 720
 QY 721 TATACCTGACCTCTACTAGCTCATGAGAGCAAAAGAGCTGTGTCCATGATGAG 780
 Db 721 TATACCTGACCTCTACTAGCTCATGAGAGCAAAAGAGCTGTGTCCATGATGAG 780
 QY 781 CATAGAGATTTGGCTGAGATTTAGAGAAAGATCTACAGATGAGAGGCGCATCTGAT 840
 Db 781 CATAGAGATTTGGCTGAGATTTAGAGAAAGATCTACAGATGAGAGGCGCATCTGAT 840
 QY 841 CCTTGTGTGAGATTTATGAGGAGAGTGGTGGTGGAGATGAAAAGATTTATGA 900
 Db 841 CCTTGTGTGAGATTTATGAGGAGAGTGGTGGTGGAGATGAAAAGATTTATGA 900
 QY 901 ATTGTCTGGCTCCATAGCTATCATCATATGATTAAGATTAAGAGATTAAGAT 960
 Db 901 ATTGTCTGGCTCCATAGCTATCATCATATGATTAAGATTAAGAGATTAAGAT 960
 QY 961 AAGTGGCTGGCATGGGCAACCAAGAGCTGTGATTAAGATTTGCAAGATTAAGCTCTT 1020
 Db 961 AAGTGGCTGGCATGGGCAACCAAGAGCTGTGATTAAGATTTGCAAGATTAAGCTCTT 1020
 QY 1021 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 Db 1021 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 QY 1081 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 Db 1081 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 QY 1081 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 Db 1081 AGAGAGCCCATCTCTATGTCACAGGCGCATGTGGATGATGATTTGATGAT 1080
 QY 1141 TTAGATGATGATTTGCTGGGCTGAGAGGCGATGATTTTCTGAGATGATTTCCGCACTG 1200
 Db 1141 TTAGATGATGATTTGCTGGGCTGAGAGGCGATGATTTTCTGAGATGATTTCCGCACTG 1200
 QY 1201 TATGCTTCTTGGCAACCAAGAGATTTGATTAAGATTTCAATCAACTCTCAAGGCAAA 1260
 Db 1201 TATGCTTCTTGGCAACCAAGAGATTTGATTAAGATTTCAATCAACTCTCAAGGCAAA 1260
 QY 1261 ACAAGGCTGAATTCGATGATTAATCATACCAAGAGATGTTTGAAGAGCTGAGGCGAG 1320
 Db 1261 ACAAGGCTGAATTCGATGATTAATCATACCAAGAGATGTTTGAAGAGCTGAGGCGAG 1320
 QY 1321 ATCTCTGAGAGCAATGAGAGCTGAACTTTAAGAAACAAGCTCTCAAAAACAGAAACAG 1380
 Db 1321 ATCTCTGAGAGCAATGAGAGCTGAACTTTAAGAAACAAGCTCTCAAAAACAGAAACAG 1380
 QY 1381 CAGGCAAGGCTGAGAGATCTGAAATTAAGAGAGAGCTGCTGTAAGATGCA 1440
 Db 1381 CAGGCAAGGCTGAGAGATCTGAAATTAAGAGAGAGCTGCTGTAAGATGCA 1440
 QY 1441 GAGAGATTAATGATGATGAG 1500
 Db 1441 GAGAGATTAATGATGATGAG 1500
 QY 1501 ATGATGATGAG 1560
 Db 1501 ATGATGATGAG 1560
 QY 1561 GCAAG 1620
 Db 1561 GCAAG 1620
 QY 1621 CAG 1680
 Db 1621 CAG 1680

QY 1681 TCAGCTTATGAGATTTCAAGAGAGAGAGATGAGAGAGATTTGAGAGAGAGAGAGATG 1740
 Db 1681 TCAGCTTATGAGATTTCAAGAGAGAGAGAGATGAGAGAGATTTGAGAGAGAGAGAGATG 1740
 QY 1741 CTGATTAAGATTAAG 1800
 Db 1741 CTGATTAAGATTAAG 1800
 QY 1801 TTTGATCTGAGAGAGAGAGATTTGATGAGGCTTTGAAACAGCTCATGATTAAGAGAGAGCTT 1860
 Db 1801 TTTGATCTGAGAGAGAGAGATTTGATGAGGCTTTGAAACAGCTCATGATTAAGAGAGAGCTT 1860
 QY 1861 CACAATGAGAGATTTGA 1878
 Db 1861 CACAATGAGAGATTTGA 1878

RESULT 3
 BT002944
 LOCUS
 DEFINITION
 Arabidopsis thaliana clone RAF1.4-93-X05 (R20243) unknown protein
 ACCESSION
 BT002944
 VERSION
 BT002944.1 GI:27754622
 KEYWORDS
 FLI CDNA.
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2162)
 Yamaoka, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
 Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
 Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayaishizaki, Y.,
 Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
 Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P.,
 Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
 Ecker, J.R. and Theologis, A.
 Arabidopsis Full Length cDNA Clones
 Unpublished
 2 (bases 1 to 2162)
 Yamaoka, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
 Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
 Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayaishizaki, Y.,
 Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
 Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P.,
 Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,
 Ecker, J.R. and Theologis, A.
 Direct Submission
 Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA (GSC) members carried out the
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayaishizaki, Y. and Shinzaki, K.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT
 JOURNAL
 COMMENT
 FEATURES
 source

The Salk, Stanford, PGC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamaoka, K., Chan, M.M.,
 Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S.,
 Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G.,
 Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M.,
 Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W.,
 Ecker, J.R. and Theologis, A.
 Yamaoka, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
 contributed equally to this work as PIs.
 Annotation based on July 2002 version of the Arabidopsis genome
 submitted to Genbank.
 Location/Qualifiers
 1..2162

```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/notes="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
5'UTR
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="A022757.1"
/db_xref="GI:27754623"
/translation="MSBRAGPMSEKKNVGVPEVQLVGLAGTRIASDDGGEM
EVSKRNKRGNTSGKTWVSONSPRAGGQOQGSVSGRNNVSGRGNGRG
IQANISGRGALSRKTDNNFVAPPVSRPFBEGMKNQAGSAHITAVEPFDVED
VDNASEBNDSDALDDSDLDSDYSDVSQSHSRKQKNKPFPSLDSLBQ
INBPQRQMHCPAQONGFALDYNLHLPLAHATKGARVYKHLAEVLEKDLQMRG
ASVIPCSEIYGQWKGJGEDEKDEIWPVPIVIMNTRLDKQDMLKMGNOELLEF
DKYALRARSYGPQGRHGSVLMFESSATGVLEAEHLRLDELAMGLDIRLAWQKSM
FGSVRLVGLATLTKODLDI FNHSGCKPLKEPLSYQEMVYKELROI SEDNOQNTY
FKRLSKONKRAKVLBSLEFIMSEKLRRTEDNRIVRQRTMQHEQNRBMQDHPF
MDSIKQIHERDAKEENFEMLQQQERAKVGGQQQINPSNDCCRRAEVSSTLEF
QEKEMEFVEREMLI KDQEKMDMKRHEBI FDLKEFDALHQLMYKHLNED
D"
1970..2162
/gene="At5g23570"
misc_difference
2146
/gene="At5g23570"
/notes="not present in genomic sequence"

ORIGIN
Query Match 100.0%; Score 1878; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGCTGCTCCATGCTTAAGAAAAGAAAGCTTCAGGCTGTTATAGGCT 60
DB 92 ATGAGTTCTAGGCTGCTCCATGCTTAAGAAAAGAAAGCTTCAGGCTGTTATAGGCT 151
QY 61 GAGGTGAACAGTTGTTCAAGGTTTGGACGAGGACGAGCTGCTTTCACAAGATGAT 120
DB 152 GAGGTGAACAGTTGTTCAAGGTTTGGACGAGGACGAGCTGCTTTCACAAGATGAT 211
QY 121 GGAGGAGAGTGGAGGCTATTTCCAGAGAAACAAGAAACAAGAAACAAGAAACAAGAA 180
DB 212 GGAGGAGAGTGGAGGCTATTTCCAGAGAAACAAGAAACAAGAAACAAGAAACAAGAA 271
QY 181 AAAAATTGGGTTTCTCAGAAATTCGAATCTCTAGAGCTTGGGTTGTCAGACGAAAGG 240
DB 272 AAAAATTGGGTTTCTCAGAAATTCGAATCTCTAGAGCTTGGGTTGTCAGACGAAAGG 331
QY 241 AGAGTAGCAACGTATCTGGAGAGAAACAATGTATCCGGAGAGGTAA CGGCAATGGT 300
DB 332 AGAGTAGCAACGTATCTGGAGAGAAACAATGTATCCGGAGAGGTAA CGGCAATGGT 391
QY 301 CGGGGCAATTAACCTAATCATATCTGTCGGGAGCGAGCGTTGACGAAAGTATGATTAAC 360
DB 392 CGGGGCAATTAACCTAATCATATCTGTCGGGAGCGAGCGTTGACGAAAGTATGATTAAC 451
QY 361 AACTTTGGCACCCCACTGTATCTCGCCCTCTTTGGAGAGAGATGAATTTGGCAG 420
DB 452 AACTTTGGCACCCCACTGTATCTCGCCCTCTTTGGAGAGAGATGAATTTGGCAG 511
QY 421 GCAAGAGAGAGTTTCTGTCAGCAACAGCTGTGACGAGATTTCTGACGTCGAGATGAT 480
DB 512 GCAAGAGAGAGTTTCTGTCAGCAACAGCTGTGACGAGATTTCTGACGTCGAGATGAT 571
```

```
QY 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 631
QY 541 CTTGCAAGTATGATTAATGATCTCGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 632 CTTGCAAGTATGATTAATGATCTCGATGATGATGATGATGATGATGATGATGATGATGAT 691
QY 601 AATAAGTGTGTTAAAAAGTTCTTTGGCACTTGGATGATGATGATGATGATGATGATGATGAT 660
DB 692 AATAAGTGTGTTAAAAAGTTCTTTGGCACTTGGATGATGATGATGATGATGATGATGATGAT 751
QY 661 GAACCAAGAGGAGGAGGATGATTCAGGCTTGCAGAAAGGAGCTTGATGATGATGATGATGAT 720
DB 752 GAACCAAGAGGAGGAGGATGATTCAGGCTTGCAGAAAGGAGCTTGATGATGATGATGATGAT 811
QY 721 TATTAACCTGACCCCTTACTAGCTCATGCGAGAGCAAAAAGAGCTTAGGCGATTAAGCTC 780
DB 812 TATTAACCTGACCCCTTACTAGCTCATGCGAGAGCAAAAAGAGCTTAGGCGATTAAGCTC 871
QY 781 CATAGGAATTTGCTTAAGTTTAAAGATCTTCAATGAGAGCGCATTTGATCATT 840
DB 872 CATAGGAATTTGCTTAAGTTTAAAGATCTTCAATGAGAGCGCATTTGATCATT 931
QY 841 CCTTGTGTGAGATTTATGAGGAGTGAAGGGTTTGGTGAAGATGAAGAAAGATTATGAA 900
DB 932 CCTTGTGTGAGATTTATGAGGAGTGAAGGGTTTGGTGAAGATGAAGAAAGATTATGAA 991
QY 901 ATTGTCTGCTCCATTAAGTGCATCATGAAATTAAGATGATGATGATGATGATGATGATGAT 960
DB 992 ATTGTCTGCTCCATTAAGTGCATCATGAAATTAAGATGATGATGATGATGATGATGATGAT 1051
QY 961 AATGCTCTGGCATGAGGCAACCAAGCTGCTGGAATTAATTGCAACAAGTATGAGCTCTT 1020
DB 1052 AATGCTCTGGCATGAGGCAACCAAGCTGCTGGAATTAATTGCAACAAGTATGAGCTCTT 1111
QY 1021 AGAGCAAGGCAATTCATAGTGCATGAGGCGCATTCGAGGATGATGATGATGATGATGATGAT 1080
DB 1112 AGAGCAAGGCAATTCATAGTGCATGAGGCGCATTCGAGGATGATGATGATGATGATGATGAT 1171
QY 1081 AGCAGTGCATCTGCTATTTTGAAGGCGCAAGCGCTTCCACCGGAGATTAGCTGATGAGGG 1140
DB 1172 AGCAGTGCATCTGCTATTTTGAAGGCGCAAGCGCTTCCACCGGAGATTAGCTGATGAGGG 1231
QY 1141 TTGATTAAGATTGCTGAGGCTCAAGAGCGCATGATTTTCTGAGAGTGTGCGCACTG 1200
DB 1232 TTGATTAAGATTGCTGAGGCTCAAGAGCGCATGATTTTCTGAGAGTGTGCGCACTG 1291
QY 1201 TATGCTTCTTGCACAGAGCAAGATCTGACATTAATCAACACTCTCAAGGCATA 1260
DB 1292 TATGCTTCTTGCACAGAGCAAGATCTGACATTAATCAACACTCTCAAGGCATA 1351
QY 1261 ACAAGGCTGAATTCAGTTGAATCATCAAGAGATGTTTAAAGAGGCTGAGGCGAG 1320
DB 1352 ACAAGGCTGAATTCAGTTGAATCATCAAGAGATGTTTAAAGAGGCTGAGGCGAG 1411
QY 1321 ATCTCTGAGCAATCAGCACTGAACTTAATTAGAACAGAGCTTCAAAACAGAACAG 1380
DB 1412 ATCTCTGAGCAATCAGCACTGAACTTAATTAGAACAGAGCTTCAAAACAGAACAG 1471
QY 1381 CAGGCGCAAGGCTTGAAGAACTCTGGAATTTATGAGGAGAAAGCTGCGTGAATCTGA 1440
DB 1472 CAGGCGCAAGGCTTGAAGAACTCTGGAATTTATGAGGAGAAAGCTGCGTGAATCTGA 1531
QY 1441 GAGGATTAATCGATCGTGAAGACAGAACTTAAGATGACATGAACAGAACAGGAGAG 1500
DB 1532 GAGGATTAATCGATCGTGAAGACAGAACTTAAGATGACATGAACAGAACAGGAGAG 1591
QY 1501 ATGATGACACAGACAGGTTTTTCAATGATTAATCAACAGATCAATGAAGAGAGAC 1560
DB 1592 ATGATGACACAGACAGGTTTTTCAATGATTAATCAACAGATCAATGAAGAGAGAC 1651
```

QY 1561 GCAGAGAGAGAGATTTCCAGATGTTGCGACGACGAGAACTGCCAAGGTTGTTGCCGAG 1620
 DB 1652 GCAGAGAGAGAGATTTCCAGATGTTGCGACGACGAGAACTGCCAAGGTTGTTGCCGAG 1711
 QY 1621 CAGCAGAGAGAGATTTCCAGATGTTGCGACGACGAGAACTGCCAAGGTTGTTGCCGAG 1680
 DB 1712 CAGCAGAGAGAGATTTCCAGATGTTGCGACGACGAGAACTGCCAAGGTTGTTGCCGAG 1771
 QY 1681 TCAGAGTTCATGAGTTTCAGAGAGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAG 1740
 DB 1772 TCAGAGTTCATGAGTTTCAGAGAGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAG 1831
 QY 1741 CTGATTAAGATTAAGAGAGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGATA 1800
 DB 1832 CTGATTAAGATTAAGAGAGAGAGATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGATA 1891
 QY 1801 TTTGATCTGAGAGAGAGATTTGATGAGGCTTTGGAACAGCTCATGTACAGATGAGCCTT 1860
 DB 1892 TTTGATCTGAGAGAGAGATTTGATGAGGCTTTGGAACAGCTCATGTACAGATGAGCCTT 1951
 QY 1861 CACATGAGAGATGATGCA 1878
 DB 1952 CACATGAGAGATGATGCA 1969

RESULT 4
 AF239719 2254 bp DNA linear PLN 02-JUN-2000
 LOCUS Arabidopsis thaliana SGS3 gene, complete cds.
 DEFINITION AF239719
 ACCESSION AF239719 GI:8164029
 VERSION AF239719.1
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana (thale cress); Eudicotyledons; core eudicots; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 2254)
 Moutreuil, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jonet, D., Lacombe, A.M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T.A. and Vaucheret, H.
 Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance Cell 101 (5), 533-542 (2000)

JOURNAL MOUTREUIL, P., BECLIN, C., ELMAYAN, T., FEUERBACH, F., GODON, C., MOREL, J.B., JONET, D., LACOMBE, A.M., NIKIC, S., PICAULT, N., REMOUE, K., SANIAL, M., VO, T.A. AND VAUCHERET, H.
 PUBLISHED 20030495
 MEDLINE 20306668
 PUBMED 10850495
 REFERENCE 2 (bases 1 to 2254)
 Beclin, C., Moutreuil, P., Vaucheret, H. and Elmayan, T.
 TITLE Direct Submision
 JOURNAL Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de Saint-Cyr, Versailles 78026, France
 FEATURES Location/Qualifiers

1. 2254
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="v"
 join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)
 /product="SGS3"
 join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)
 /codon_start=1
 /product="SGS3"
 /protein_id="AF239719.1"
 /db_xref="GI:8164030"
 /translation="MSRAGPWSKKNVQGYRPREVDVVOGLAATRLASSODDGGEM
 EVISKKNKPKNTSGKTWVSQNSNPRAWGQOQGRSNVSGKNVSGRNGGK
 IOANISGRGRALSRKYDNNFVAPVPVSPPLGGNNWARGSAQHTAVQFPEVED
 VDNASEENDSDALDDSDLDASDDYDQSHSGSRKQKWFKEFGSLDSISIBQ
 INEPORHWCACQNGPGALIDYNNLPLLAHRTGARGVRLHLEAVELEKIDQMG
 ASVTFCGEIYQGMKGLGDEKDYETVMPMVTIMNTRLDXNDKNGKMGQELLEYF
 DKTEALRAHSHYGPQGHKGMVLMFESSATGTLEBARHRLAENGKLDRIAMGKRSM
 FSGVROLVGFPLATKQDLIDFNQHSQGRILKFLKSKYQEMVVELROIISDNOQNT

ORIGIN

Query Match 51.3%; Score 963; DB 8; Length 2254;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTTACGGCTGGTCCAAATGCTTAGGAGAAAGAGCTTCAGGGTGTATAGGCT 60
 1 ATGAGTTTACGGCTGGTCCAAATGCTTAGGAGAAAGAGCTTCAGGGTGTATAGGCT 60
 QY 61 GAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGACTGCTTCTTCAAGATGAT 120
 DB 61 GAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGACTGCTTCTTCAAGATGAT 120
 QY 121 GGAGGAGAGTGGAGAGTCAATTCAGAGAAACAAGCAACAGAGAAACATTTTGA 180
 DB 121 GGAGGAGAGTGGAGAGTCAATTCAGAGAAACAAGCAACAGAGAAACATTTTGA 180
 QY 181 AAACTTGGGTTTCTCAAGATTCGAATCTCTGAGCTTGGGTTGTCAGCAGCAAGG 240
 DB 181 AAACTTGGGTTTCTCAAGATTCGAATCTCTGAGCTTGGGTTGTCAGCAGCAAGG 240
 QY 241 AGAGGTAGCAACGTATCTGAGAGAGAAACATGATTCGGGAGAGTAAACGCAATGCT 300
 DB 241 AGAGGTAGCAACGTATCTGAGAGAGAAACATGATTCGGGAGAGTAAACGCAATGCT 300
 QY 301 CGGGGCAATTCAGCTTAACATATCTGTTGGGAGCAGAGGCTTGAGCAGAAATGATAC 360
 DB 301 CGGGGCAATTCAGCTTAACATATCTGTTGGGAGCAGAGGCTTGAGCAGAAATGATAC 360
 QY 361 AACTTTGGGCAACCCCACTGATCTGACCTCTTTGGAAGAGATGAAATTTGCGAG 420
 DB 361 AACTTTGGGCAACCCCACTGATCTGACCTCTTTGGAAGAGATGAAATTTGCGAG 420
 QY 421 GCAGAGAGAGTTCCTGCTCAGACACAGCTGTGCAGAGATTCTGACCTGAGAGATAT 480
 DB 421 GCAGAGAGAGTTCCTGCTCAGACACAGCTGTGCAGAGATTCTGACCTGAGAGATAT 480
 QY 481 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 DB 481 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 QY 541 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 DB 541 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 QY 541 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 DB 541 GTGATATGCTTCTGAGAGAAAGAAATGATTCGATGCTTGGAGATTCGATGAGCAG 540
 QY 601 AATAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGTGATGAGCAGATTAAT 660
 DB 601 AATAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGTGATGAGCAGATTAAT 660
 QY 661 GAACCAAGAGAGAGAGATGCTTCCAGCTTTCAGAAACGAGCTGATGATGAGTGG 720
 DB 661 GAACCAAGAGAGAGAGATGCTTCCAGCTTTCAGAAACGAGCTGATGATGAGTGG 720
 QY 721 TATACCTGACCCCTTACTAGCTCATGCGAGAGCAAAAAGAGCTTAGGAGTTAAGCTC 780
 DB 721 TATACCTGACCCCTTACTAGCTCATGCGAGAGCAAAAAGAGCTTAGGAGTTAAGCTC 780
 QY 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTTCAGATGAGAGGCGCATCTGCAT 840
 DB 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTTCAGATGAGAGGCGCATCTGCAT 840
 QY 841 CTTTGGTGAAGATTATGAGGAGAGTGAAGAGGTTTGGTGAAGATGAAGATTAAGAA 900
 DB 841 CTTTGGTGAAGATTATGAGGAGAGTGAAGAGGTTTGGTGAAGATGAAGATTAAGAA 900
 QY 901 ATTGTGAGGCTCAATGCTCATCATGAATATCTAGACTGATTAAGAGATTAAGAT 960
 DB 901 ATTGTGAGGCTCAATGCTCATCATGAATATCTAGACTGATTAAGAGATTAAGAT 960

QY	961	AAG	963	
Db	961	AAG	963	
RESULT 5				
AX078760				
LOCUS	AX078760	3275 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0105951.			
ACCESSION	AX078760			
VERSION	AX078760.1	GI:13158379		
KEYWORDS				
SOURCE				
ORGANISM	Arabidopsis thaliana (thale cress)			
	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE				
AUTHORS	1			
TITLE	Beclin, C., Elmayer, T. and Vaucheret, H.			
JOURNAL	Novel sg83 plant gene and use thereof			
	Patent: WO 0105951-A 1 25-JAN-2001;			
	AVENTIS CROPS/SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE			
	AGRONOMIQUE (FR)			
FEATURES				
source	Location/Qualifiers			
	1..3275			
	/organism="Arabidopsis thaliana"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:3702"			
	693..715			
	/note="P356AD;"			
	complement(2926..2952)			
	/note="P356Y;"			
ORIGIN				

Db	1176	GTGATATATAGCTTCTGAGGAAGAATGATTCGATGCTTTGGATGATTTCTGATGAGCAG	1235
Qy	541	CTTGCAAGTGAATGATTTATGACTCGGATGTGATCAAAAAGCCATTGATCAAGAACAG	600
Db	1236	CTTGCAAGTGAATGATTTATGACTCGGATGTGATCAAAAAGCCATTGATCAAGAACAG	1295
Qy	601	AATAAGTGTTCAAAAAGTCTTTGGCAGCTTGATAGCTTGTGATCGAGCATATAAT	660
Db	1296	AATAAGTGTTCAAAAAGTCTTTGGCAGCTTGATAGCTTGTGATCGAGCATATAAT	1355
Qy	661	GAACCAAGAGGCAGTGGCATTGTCCAGCTTTGTCAAAACGGACTTGGTCCATTGATTTGG	720
Db	1356	GAACCAAGAGGCAGTGGCATTGTCCAGCTTTGTCAAAACGGACTTGGTCCATTGATTTGG	1415
Qy	721	TATACCTGCACCCCTTACTAGCTCATGCGAGCAAAAAGAGCTTAGCGAGTTAAGCTC	780
Db	1416	TATACCTGCACCCCTTACTAGCTCATGCGAGCAAAAAGAGCTTAGCGAGTTAAGCTC	1475
Qy	781	CATAGAGAATTTGCTGAAGTTTGA AAAAGGATCTACAGATGAGAGCGCATCTGTCAAT	840
Db	1476	CATAGAGAATTTGCTGAAGTTTGA AAAAGGATCTACAGATGAGAGCGCATCTGTCAAT	1535
Qy	841	CCTTGCTGTAGATTTATGAGCGAGTGAAGGCTTTGGGTAGAGATGAAAAGGATTAAGAA	900
Db	1536	CCTTGCTGTAGATTTATGAGCGAGTGAAGGCTTTGGGTAGAGATGAAAAGGATTAAGAA	1595
Qy	901	ATTGTCTGGCTCCCAATGATCATCATGATACTAGACTGATTAAGAGAGATTAACGAT	960
Db	1596	ATTGTCTGGCTCCCAATGATCATCATGATACTAGACTGATTAAGAGAGATTAACGAT	1655
Qy	961	AAG 963	
Db	1656	AAG 1658	

Query Match	51.3%	Score 963	DB	Length 3275
Best Local Similarity	100.0%	Pred. No. 0		
Matches	963	Conservative 0	Mismatches 0	Indels 0
			Gaps 0	
QY	1	ATGAGTTCTAGAGGCTGCTCCATATGCTTAAGGAAAAAGAAAGTTTCAGAGGTGTTATAGGCTT	60	
Db	696	ATGAGTTCTAGAGGCTGCTCCATATGCTTAAGGAAAAAGAAAGTTTCAGAGGTGTTATAGGCTT	755	
QY	61	GAGGTTGAACAGTTGGTCTCAAGGTTTGGCAGAGGACGACCTGGCTTCTTCCAAAGATAT	120	
Db	756	GAGGTTGAACAGTTGGTCTCAAGGTTTGGCAGAGGACGACCTGGCTTCTTCCAAAGATAT	815	
QY	121	GGAGAGAGTGGGAGGTCATTCTCCAGAGAAACAAGAACCAACGAGAAACATCTTGGGA	180	
Db	816	GGAGAGAGTGGGAGGTCATTCTCCAGAGAAACAAGAACCAACGAGAAACATCTTGGGA	875	
QY	181	AAAACTTGGGTTTCTCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTGACGACGACGG	240	
Db	876	AAAACTTGGGTTTCTCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTGACACGACGACGG	935	
QY	241	AGAGTAGCAACGATCTCGGAGAGAGAAACAATGATATCGGAGAGAGGTAAACGGCAATGCT	300	
Db	936	AGAGTAGCAACGATCTCGGAGAGAGAAACAATGATATCGGAGAGAGGTAAACGGCAATGCT	995	
QY	301	CGGGGCATTCAAGCTAACATATCTGTCGGGACGAGCGTTGAGCAGAAATATGATTAAC	360	
Db	996	CGGGGCATTCAAGCTAACATATCTGTCGGGACGAGCGTTGAGCAGAAATATGATTAAC	1055S	
QY	361	AACCTTGGGACACCCCAACCTGTATCTCGCCCTCCTTTGAAAGAGATGGAATTGGAG	420	
Db	1056	AACCTTGGGACACCCCAACCTGTATCTCGCCCTCCTTTGAAAGAGATGGAATTGGAG	1115S	
QY	421	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGCGAGAGTTTCTGACGTGAGAGATAT	480	
Db	1116	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGCGAGAGTTTCTGACGTGAGAGATAT	1175S	
QY	481	GTGATATATGTTCTGAGAGAAAGATATATTCGATGCTTTGGATGATTTCTGATGACAC	540	

LOCUS	AB025633	81365 bp	DNA	linear	PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, F1 clone:MOM1.				
ACCESSION	AB025633 BA000015				
VERSION	AB025633.2 GI:10178221				
KEYWORDS	.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
REFERENCE	1 (sites)				
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones				
JOURNAL	DNA Res. 7 (1), 31-63 (2000)				
MEDLINE	20181125				
PUBMED	10718197				
REFERENCE	2 (bases 1 to 81365)				
AUTHORS	Nakamura,Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yakamun@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
COMMENT	On Sep 15, 2000 this sequence version replaced gi:4589439. Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/xaos/cgi-bin/asg.graph.cgi?c=MOM1 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/),				

GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>), NetGene2 (S.-M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://genome1.zoology.berkeley.edu/cgi-bin/sp.cgi>). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K19M13 and the 3' clone is M9011.

FEATURES

Source

1. .81365
/arabidomaia thaliana"

```
/organism="Arabidopsis thaliana"
```

/mol_type="genomic DNA"

```
/strain="Columbia"
```

/db_xref="taxon:3702"

```
/chromosome="5"
```

```
clone="MQM1"
```

```
/clone lib="Mitsui P1"
```

```

/codon_start=1
/evidence=not experimental
/protein_id="BA97243.1"
/db_xref="GI:8609702"
/translation="MAPPPGQRTAAVAEAGADGAPPGQQRGFGSTISGIVIAVNYE
ASRFPSPKQPMQDPAVSQMLNLFPHKGSILDMWFLSIOEKNDGDNALYWHETN
IPYAVWPBSIRKSLIYPSSEITLONNNGSYAHIPARSGPFDIPDPBPQPLNSR
THAVATYPRKQKNNKSLSPKSDSESEPEVEKDKSDKEEVPVEMSLMKPN
VTINLVDDPTFOADILVEPTGNYPYITFENREILRDKEIPVNETVSELPLNEISF
ISMWKOLFOQUDOSFONORSXGSMIDGSDDELKRYPLSGNPVLIGITMPSMLHSVP
DFIAFKNDIQFNKNSMBGLSAKSVLNPICQFVFLYILNDTSMILASGVVC
IEFWKIGKMRLEWDRSGMIPLRFHDSRYASNKEVDLIAIKELSYVLLVLVLG
SISYLAVERHKSIVSLSLISLSCVYMGFIMMCPQPLFNLYKLKSYAHLPMRWMTYKF
LNTIIDLFAFVIMKPIHLRLSVFRDVLFLYLQRMWYVVDKTRVNEFGFGEDET
AEKRLITEKEBEDKKTN"
complement(join(29439..29576,29661..29885,29985..30268,
30341..30493,30579..30722,30956..31484))
/note="gene_id:MOM1.15

```

```

Query Match      51.3%; Score 963; DB 8; Length 81365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGAGTTCTAGGGCTGCTCCAAATGTCTAAGAAAGAAAGCCTTCAAGGTGTATAGCCT 60
DB 22672 ATGAGTTCTAGGGCTGCTCCAAATGTCTAAGAAAGAAAGCCTTCAAGGTGTATAGCCT 22731
QY 61 GAGGTGAACAGTTGGTTCAGAGTTGGCGAGGACGAGACTGGCTTCTTACAGATGAT 120
DB 22732 GAGGTGAACAGTTGGTTCAGAGTTGGCGAGGACGAGACTGGCTTCTTACAGATGAT 22791
QY 121 GGAGGAGTGGGAGGCTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 22792 GGAGGAGTGGGAGGCTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 22851
QY 181 AAAAAGTTGGTTTCTCAGAAATTCCTCTAGAGCTTGGGCTGTCAGCAGCAAGG 240
DB 22852 AAAAAGTTGGTTTCTCAGAAATTCCTCTAGAGCTTGGGCTGTCAGCAGCAAGG 22911
QY 241 AAGAGTAGCAACGATATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 22912 AAGAGTAGCAACGATATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 22971
QY 301 CGGGGATTCAGACTAATATCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 22972 CGGGGATTCAGACTAATATCTGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23031
QY 361 AACTTTGGGACCCGCCCTGTATCTCGCCCTCTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 23032 AACTTTGGGACCCGCCCTGTATCTCGCCCTCTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 23091
QY 421 GCAGAGAGAGTTCTGCTCAGCAGCAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480
DB 23092 GCAGAGAGAGTTCTGCTCAGCAGCAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 23151
QY 481 GTGAGTAGATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 23152 GTGAGTAGATGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23211
QY 541 CTTGCAAGTAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 23212 CTTGCAAGTAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 23271
QY 601 AATAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGCTTGTGATCGAGCAGATTAAT 660
DB 23272 AATAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGCTTGTGATCGAGCAGATTAAT 23331
QY 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 23332 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23391
QY 721 TATAACCTGACCCCTTACTAGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TATAACCTGACCCCTTACTAGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

```

DB 23392 TATAACCTGACCCCTTACTAGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23451
QY 781 CATAGAGATTTGGCTGAGTTTATAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 23452 CATAGAGATTTGGCTGAGTTTATAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAGAGAG 23511
QY 841 CTTGTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 23512 CTTGTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23571
QY 901 ATTGTGCTGCTTCAATGCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 960
DB 23572 ATTGTGCTGCTTCAATGCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 23631
QY 961 AAG 963
DB 23632 AAG 23634

```

```

RESULT 7
ATH528171
LOCUS 650 bp DNA linear PLN 28-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone
157C08.
ACCESSION AJ528171
VERSION AJ528171.1 GI:26796431
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Sanson, F.,
Chauvin, S., Bechoff, N., Cruaud, C., Derose, R., Pelletier, G.,
Lepoint, L., Caboche, M. and Lecharny, A.
T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)
JOURNAL MEDLINE 22363535
PUBMED 12446565
REFERENCE 2 (bases 1 to 650)
AUTHORS Balzergue, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzergue S., IMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplance' (http://www.genoplance.com and
http://genoplance-info.infobiogen.fr).
FEATURES
source
1..650
Location/Qualifiers
1..650
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Massillaewikijs"
/db_xref="taxon:3702"
/clone="157C08"
/clone_1ib="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
1..650
/note="T-DNA flanking sequence
left border"
ORIGIN
Query Match 12.7%; Score 239; DB 8; Length 650;
Best Local Similarity 99.7%; Pred. No. 9,9e-131;
Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```


QY	962	AGTGGCTCGGATGGGCAACCAAGCTGCTGGAAATCTTCGACAAAGTATGAGGCTCTTA	1021
Db	303	AGTGGCTCGGATGGGCAACCAAGCTGCTGGAAATCTTCGACAAAGTATGAGGCTCTTA	362
QY	1022	GAGCAACGCATTCCTATATGATCCACAGGGCCATCGTGGATGATGTTCGATGTTTAA	1081
Db	363	GAGCAACGCATTCCTATATGATCCACAGGGCCATCGTGGATGATGTTCGATGTTTAA	422
QY	1082	GCAATGCCATCGGCTATTTGGAGGCGCAACCCCTCCACCGGAGATTAGCTGAGATGGGGT	1141
Db	423	GCAATGCCATCGGCTATTTGGAGGCGCAACCCCTCCACCGGAGATTAGCTGAGATGGGGT	482
QY	1142	TAGATAGAAATTCCTCGGGGTCAAAAGCCGATATGTTTTCGGAGGATTCGGCCAACTGT	1201
Db	483	TAGATAGAAATTCCTCGGGGTCAAAAGCCGATATGTTTTCGGAGGATTCGGCCAACTGT	542
QY	1202	ATGGCTTCCTTCAACGAAGCAAGATCTGACATATTCATCAACACTCT	1251
Db	543	ATGGCTTCCTTCAACGAAGCAAGATCTGACATATTCATCAACACTCT	592

RESULT 8			
AC121725/c			
LOCUS	178928 bp	DNA	/ linear
DEFINITION	Rattus norvegicus clone CH230-328r2,	***	SEQUENCING IN PROGRESS
***	3 unordered pieces.		

ACCESSION	AC121725	GI:23804926	
VERSION	AC121725.3		
KEYWORDS	HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus		
ORGANISM	Rattus norvegicus	Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eumammalia; Eutheria; Rodentia; Muridae;	

REFERENCE
AUTHORS

(Bases 1 to 178928)

Mizny, D., Marie, Metzker, M., Lee, Aramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguitano, D., Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baidun, D., Bandaranaike, D., Barber, M., Barnstead, M., Bernhardt, F., Bielawski, K., Blair, J., Blanchburg, K., Bluth, P., Brown, M., Bryant, N., Buhay, C., Burck, P., Buttell, K., Caldoro, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, L., Cockrell, R., Cox, C., Coyle, M., Crete, A., De Souza, L., Delgado, O., Demson, S., Detamo, C., Ding, Y., Dinsh, H., Diya, K., Draper, H., Dugan-Kocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, I., Garza, M., Gebregorjis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, Gunnarane, P., Healand, W., Hamll, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Hines, S., Hladson, S.L., Hodgson, A., Hogues, M., Hernandez, R., Hines, S., Hladon, S.L., Hummel, J., Idlebird, D., Jackson, A., Hollins, B., Howells, S., Huixy, S., Hume, J., Idelbird, D., Joelvet, A., Jacobson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Johnson, K., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Karpathy, S., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Kowals, C., Kraft, C.L., London, P., Longacre, S., Lopez, J., Liu, J., Liu, W., Liu, Y., Louissegod, H., Lorado, R.J., Lu, X., Ma, J., Lorenzshew, L., Louissegod, H., Lorado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Manjivney, S., McLeod, M.P., McNelly, T.Z., Meentemey, J., Milosavljevic, A., Miner, G., Minja, E., Mundada, M., Murphy, M., Morgan, M., Morris, K., Morris, S., Mundada, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakilemele, O., Okunodu, G., Olarnmuagaon, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plapper, F., Polndexter, A., Popovic, D., Primus, E., Fu, L., L., Puzoz, M., Quiroz, J., Rachlin, E., Reeves, K., Reiser, M.A., Religh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojs, A., Rose, M., Rose, R., Ruiz, S.J.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Sanders, W., Savery, G., Scheer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaj, D., Sneet, A., Sodegriem, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinle, M., Thomas, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaana, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlezyk, R., Woodeen, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiser, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinsthausen, G., and Gibbs, R.A.

Direct Submission

Unpublished

2. (bases 1 to 178928)

Worley, K.C.

Direct Submission

Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3. (bases 1 to 178928)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT

On Oct 11, 2002 this sequence version replaced gi:21909375. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atl1/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYJD
Center clone name: CH230-32812
----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 143032 bases at least Q40
Consensus quality: 146972 bases at least Q20
Consensus quality: 149332 bases at least Q20
Estimated insert size: 145820; sum-of-contrigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contrigs estimation

```

```

*-----*
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
*
* NOTE: This is a "working draft" sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1      175436: contig of 175436 bp in length
*      *      175536: gap of unknown length
*      *      175537: contig of 1043 bp in length
*      *      175580: gap of unknown length
*      *      176679: gap of unknown length
*      *      176680: contig of 2249 bp in length.

```

FEATURES
source

```

I. .I/0520
/organism="Rattus norvegicus"

```



```

misc_feature      /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-32812"
                  1..1169
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      27453..29430
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      77608..79432
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      99609..100719
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      /note="wgs_end_extension
clone_end:Sp6"
misc_feature      complement(118199..119034)
                  /note="clone_boundary
clone_end:Sp6
misc_feature      end:sequence:RXA0N49TV"
                  125929..127858
                  /note="wgs_contig"
misc_feature      156410..157915
                  /note="wgs_contig"
misc_feature      complement(172023..172895)
                  /note="clone_boundary
clone_end:T7
misc_feature      site:MboI
                  end:sequence:RXA0N49TV"
                  173824..175436
                  /note="wgs_end_extension
clone_end:T7"

```

```

ORIGIN
Query Match      1.2%; Score 22; DB 2; Length 178928;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 492 TTCTGAGAGAGATGATTC 513
Db 112510 TTCTGAGAGAGATGATTC 112489

```

```

RESULT 9
AL928719/c      197909 bp      DNA      linear      ROD 28-JAN-2003
LOCUS           Mouse DNA sequence from clone RP23-419G21 on chromosome 2, complete
DEFINITION      sequence.
ACCESSION       AL928719
VERSION         AL928719.6
KEYWORDS        HTG.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 197909)
REFERENCE       1 (bases 1 to 197909)
AUTHORS        Leongamornlert, D.
TITLE           Direct Submission
JOURNAL         Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Nov 3, 2002 this sequence version replaced gi:24395353.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
COMMENT         ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL, Sw., SWISSPROT, Tr., TREMBL, Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C/elegans/wormpep> RP23-419G21 is from the RPI-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6.

```

FEATURES
source
Location/Qualifiers
1..197909
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-419G21"
/clone_11b="RPI-23"

```

```

ORIGIN
Query Match      1.2%; Score 22; DB 10; Length 197909;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGACTCAAGAGCCATGCA 588
Db 34650 TGTGACTCAAGAGCCATGCA 34629

```

```

RESULT 10
AC078933      214166 bp      DNA      linear      HTG 11-AUG-2000
LOCUS           Mus musculus chromosome 5 clone RP23-201E13 strain C57BL6/J,
DEFINITION      WORKING DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION       AC078933
VERSION         AC078933.1
KEYWORDS        HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214166)
REFERENCE       1 (bases 1 to 214166)
AUTHORS        Bouffard, G.G., Dietrich, N.L., Bagle, W.O., Gupta, J., Ho, S.-L.,
Maestrian, S.D., McCloskey, J.C., McDowell, J., Ojodu, M.A., Pearson, R.,
Stanciripop, S., Summers, T.J., Thomas, V.W., Thomas, P.J.,
Tlionson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., and Green, E.D.
TITLE           NISC Mouse Sequencing Initiative
JOURNAL         Unpublished
2 (bases 1 to 214166)
REFERENCE       Green, E.D.
AUTHORS        Direct Submission
TITLE           Submitted (11-AUG-2000) NIH Intramural Sequencing Center, 8717
JOURNAL         Groveomont Circle, Gaithersburg, MD 20877, USA

```

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_mouse@nih.gov

```

----- Project Information
Center project name: vd
Center clone name: 201E13
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.99019
Consensus quality: 204605 bases at least Q40
Consensus quality: 207891 bases at least Q30
Consensus quality: 209797 bases at least Q20
Insert size: 218000; agarose-gel
Insert size: 237000; pulse-field-gel
Insert size: 212586; sum-of-contigs
Quality coverage: 4.82x in Q20 bases; agarose-gel
Quality coverage: 4.43x in Q20 bases; pulse-field-gel
Quality coverage: 4.94x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3207: contig of 3207 bp in length
3208 3307: gap of unknown length
3308 7622: contig of 4315 bp in length
7623 7722: gap of unknown length
7723 13087: contig of 5365 bp in length
13088 13187: gap of unknown length
13188 20448: contig of 7261 bp in length
20449 20549: gap of unknown length
20549 26559: contig of 6111 bp in length
26560 26760: gap of unknown length
26760 35349: contig of 8590 bp in length
35350 35449: gap of unknown length
35450 43812: contig of 8363 bp in length
43813 52486: contig of 8574 bp in length
52487 52586: gap of unknown length
52587 62100: contig of 9514 bp in length
62101 62200: gap of unknown length
62201 77345: contig of 15145 bp in length
77346 77445: gap of unknown length
77446 87428: contig of 9983 bp in length
87429 87529: gap of unknown length
87529 100606: contig of 13078 bp in length
100607 100707: gap of unknown length
100707 117904: contig of 17198 bp in length
117905 118005: gap of unknown length
118005 135715: contig of 17711 bp in length
135716 135815: gap of unknown length
135816 153408: contig of 17593 bp in length
153409 153509: gap of unknown length
153509 181134: contig of 27526 bp in length
181135 214186: gap of unknown length
181135 214186: contig of 33052 bp in length.
-----
FEATURES
Source
1..214186
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone_id="RP23-201E13"
/clone_lib="RPCI mouse BAC library 23"
1..3207
/misc_feature
/note="assembly_fragment"
3308..7622
/misc_feature
/note="assembly_fragment"
7723..13087
/misc_feature
/note="assembly_fragment"

```

```

misc_feature 13188..20448
/note="assembly_fragment"
misc_feature 20549..26659
/note="assembly_fragment"
misc_feature 26760..35349
/note="assembly_fragment"
misc_feature 35450..43812
/note="assembly_fragment"
misc_feature 43913..52486
/note="assembly_fragment"
misc_feature 52587..62100
/note="assembly_fragment"
misc_feature 62201..77345
/note="assembly_fragment"
misc_feature 77446..87428
/note="assembly_fragment"
misc_feature 87529..100606
/note="assembly_fragment"
misc_feature 100707..117904
/note="assembly_fragment"
misc_feature 118005..135715
/note="assembly_fragment"
misc_feature 135816..153408
/note="assembly_fragment"
misc_feature 153509..181034
/note="assembly_fragment"
misc_feature 181135..214186
/note="assembly_fragment"
-----
ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 214186;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 567 TGTGAGTCAAAAGAGCCATGGA 588
Db 99680 TGTGAGTCAAAAGAGCCATGGA 99701
-----
RESULT 11
AC079182/c 218774 bp DNA linear HTG 23-AUG-2000
LOCUS Mus musculus chromosome 5 clone RP23-203F6 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION AC079182
VERSION AC079182.1 GI:9886001
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 218774)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Idol,J., Lee-Jin,S.-O., Legaspi,R., Lim,M., Maduro,Q.L.,
Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
Pearson,R., Stantitrop,S., Sumner,T.J., Thomas,J.W., Thomas,P.J.,
Tongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Weberby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 218774)
Green,E.D.
Direct Submission
Submitted (23-AUG-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
COMMENT

```

```

Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgr1.nih.gov
-----
Project Information
Center project name: ve
Center clone name: 203P06
-----
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Assembly quality: 208040 bases at least Q40
Consensus quality: 21445 bases at least Q30
Consensus quality: 213463 bases at least Q20
Insert size: 227000; agarose-tp
Insert size: 242000; pulse-field-gel
Insert size: 217274; sum-of-ctdigs
Quality coverage: 4.70x in Q20 bases; agarose-tp
Quality coverage: 4.40x in Q20 bases; pulse-field-gel
Quality coverage: 4.91x in Q20 bases; sum-of-ctdigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
1 2001: contig of 2001 bp in length
2 2002 2101: gap of unknown length
3 2102 4870: contig of 2769 bp in length
4 4871 4970: gap of unknown length
5 4971 8469: contig of 3499 bp in length
6 8470 8569: gap of unknown length
7 8570 14232: contig of 5663 bp in length
8 14233 14332: gap of unknown length
9 14333 21631: contig of 7299 bp in length
10 21632 21731: gap of unknown length
11 21732 29661: contig of 7930 bp in length
12 29662 29761: gap of unknown length
13 29762 38142: contig of 8381 bp in length
14 38143 38242: gap of unknown length
15 38243 47730: contig of 9488 bp in length
16 47731 47830: gap of unknown length
17 47831 55944: contig of 8114 bp in length
18 55945 56044: gap of unknown length
19 56045 70548: contig of 14504 bp in length
20 70549 70648: gap of unknown length
21 70649 85515: contig of 14867 bp in length
22 85516 85615: gap of unknown length
23 85616 103358: contig of 17743 bp in length
24 103359 103458: gap of unknown length
25 103459 125084: contig of 21626 bp in length
26 125085 125184: gap of unknown length
27 125185 153356: contig of 28172 bp in length
28 153357 153456: gap of unknown length
29 153457 184845: contig of 31389 bp in length
30 184846 184945: gap of unknown length
31 184946 218774: contig of 33829 bp in length.
-----
Location/Qualifiers
1. 218774
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone_lib="RP23-203P6"
/clone_lib="PPCI mouse BAC library 23"
1. 2001
/notes="assembly_fragment"
2102. 4870
/notes="assembly_fragment"
4971. 8469
/notes="assembly_fragment"
misc_feature
/notes="assembly_fragment"

```

```

misc_feature 8570. 14232
/notes="assembly_fragment"
misc_feature 1433. 21631
/notes="assembly_fragment"
misc_feature 21732. 29661
/notes="assembly_fragment"
misc_feature 29762. 38142
/notes="assembly_fragment"
misc_feature 38243. 47730
/notes="assembly_fragment"
misc_feature 47831. 55944
/notes="assembly_fragment"
misc_feature 56045. 70548
/notes="assembly_fragment"
misc_feature 70649. 85515
/notes="assembly_fragment"
misc_feature 85616. 103358
/notes="assembly_fragment"
misc_feature 103459. 125084
/notes="assembly_fragment"
misc_feature 125185. 153356
/notes="assembly_fragment"
misc_feature 153457. 184845
/notes="assembly_fragment"
misc_feature 184946. 218774
/notes="assembly_fragment"
-----
ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 218774;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 567 TGTGACTCAAAAGAGCCATGCA 588
Db 76509 TGTGACTCAAAAGAGCCATGCA 76488
-----
RESULT 12
AC115306/c 236508 bp DNA linear HTG 23-NOV-2002
LOCUS Rattus norvegicus clone CH230-11F1, WORKING DRAFT SEQUENCE, 3
DEFINITION
unordered pieces.
ACCESSION
AC115306.4 GI:25188797
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus
ORGANISM
Rattus norvegicus
Mammalia; Eutheria; Chordata; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 236508)
Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bunay, C., Butch, P., Butrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Casas, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapeer, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnatane, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,

```

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowals, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martineau, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munda, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, G., Olarnunagsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plummer, F., Poindexter, A., Popovic, D., Primm, E., Pu, L., Puzos, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Smetly, J., Snavarsky, A., Song, X.-Z., Sorelle, R., Sosa, J., Sneed, A., Sodergren, E., Sutton, A., Syarik, A., Taber, P., Taylor, C., Steimle, M., Strong, R., Thomas, S., Tingey, A., Treitz, Z., Umani, K., Taylor, T., Thomas, N., Thomas, S., Waldron, J., Walker, B., Wang, J., Vallas, R., Vera, V., Villaseca, D., Waldron, J., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE
Unpublished
REFERENCE
2 (bases 1 to 236508)
AUTHORS
Worley, K.C.
JOURNAL
Submitted (17-Mar-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
Rat Genome Sequencing Consortium.
AUTHORS
Submitted (23-Nov-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 23, 2002 this sequence version replaced g1:22772519. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBAS
Center clone name: CH230-11F1
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 219730 bases at least Q40
Consensus quality: 222791 bases at least Q30
Consensus quality: 224323 bases at least Q20
Estimated insert size: 225524; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

***** NOTE: Estimated insert size may differ from sequence length

FEATURES
source
1. 236508 "Rattus norvegicus"
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-11F1"
1. 1641
/note="wgs contig"
155976..157691
/note="wgs contig"
233371..235181
/note="wgs contig"
235232..236508
/note="wgs contig"

ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 236508;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
161 AACGAGAAACCTTGGAA 182
Db 162467 AACGAGAAACCTTGGAA 162446

RESULT 13
AC140364 280016 bp DNA linear HTG 05-NOV-2003
LOCUS
DEFINITION
SEQUENCE, 8 unordered pieces.
AC140364
AC140364.2 GI:38176012
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 280016)

REFERENCE
1 Wilson, R.K.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 280016)
AUTHORS
McPherson, J.D. and Waterston, R.H.
TITLE
Submitted (23-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
JOURNAL
3 (bases 1 to 280016)
REFERENCE
Wilson, R.K.
Submitted (05-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
TITLE
On Nov 5, 2003 this sequence version replaced g1:28475622.
JOURNAL

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
----- Project Information -----

Center project name: M.BB0075K05

----- Summary Statistics -----

Sequencing vector: M13; 0%
Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 278028 bases at least Q40
Consensus quality: 278887 bases at least Q30
Consensus quality: 279492 bases at least Q20

NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 3866: contig of 2564 bp in length
* 3867 3966: gap of unknown length
* 3967 24768: contig of 20802 bp in length
* 24769 24868: gap of unknown length
* 24869 48433: contig of 23565 bp in length
* 48434 48533: gap of unknown length
* 48534 85803: contig of 37270 bp in length
* 85804 85903: gap of unknown length
* 85904 127685: contig of 41782 bp in length
* 127686 127785: gap of unknown length
* 127786 198220: contig of 70435 bp in length
* 198221 198320: gap of unknown length
* 198321 280016: contig of 81696 bp in length.
Location/Qualifiers

1. .280016
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-75K5"

misc_feature /note="assembly_name:Contig20"
misc_feature 1303..3866
/note="assembly_name:Contig22"
misc_feature 3867..24768
/note="assembly_name:Contig23"
misc_feature 24869..48433
/note="assembly_name:Contig24"
misc_feature 48534..85803
/note="assembly_name:Contig25"
misc_feature 85904..127685
/note="assembly_name:Contig26"
misc_feature 127786..198220
/note="assembly_name:Contig27"
misc_feature 198321..280016
/note="assembly_name:Contig28"

ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 280016;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 TGTGAGTCAGAGAGCCATGGA 588
DB 58633 TGTGAGTCAGAGAGCCATGGA 58654

RESULT 14
AC073813 AC073813 304407 bp DNA linear HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-69B1, WORKING DRAFT SEQUENCE, 58 unordered

ACCESSION AC073813
VERSION AC073813.1 GI:8910430
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 304407)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 304407)
DOE Joint Genome Institute.
Direct Submission
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1763393
Center Clone name: RPCI-23_69B1

Summary Statistics
Consensus quality: 257140 bases at least Q40
Consensus quality: 282847 bases at least Q30
Consensus quality: 288639 bases at least Q20
Estimated insert size: 272860; agarose-fp estimation
Estimated coverage: 9.46 in Q20 bases; agarose-fp estimation
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently consists of 58 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 1167: contig of 1167 bp in length
* 1168 1267: gap of unknown length
* 1268 2281: contig of 1014 bp in length
* 2282 2381: gap of unknown length
* 2382 3767: contig of 1366 bp in length
* 3768 3867: gap of unknown length
* 3868 5372: contig of 1505 bp in length
* 5373 5472: gap of unknown length
* 5473 6484: contig of 1022 bp in length
* 6485 6594: gap of unknown length
* 6595 7909: contig of 1315 bp in length
* 7910 8009: gap of unknown length
* 8010 9416: contig of 1407 bp in length
* 9417 9516: gap of unknown length
* 9517 10776: contig of 1260 bp in length
* 10777 10876: gap of unknown length
* 10877 12279: contig of 1403 bp in length
* 12280 12379: gap of unknown length
* 12380 13920: contig of 1541 bp in length
* 13921 14020: gap of unknown length
* 14021 15417: contig of 1397 bp in length
* 15418 15517: gap of unknown length
* 15518 16897: contig of 1380 bp in length
* 16898 16997: gap of unknown length
* 16998 18940: contig of 1943 bp in length
* 18941 18941: gap of unknown length
* 18942 20880: contig of 1850 bp in length
* 20881 20990: gap of unknown length
* 20991 22965: contig of 1975 bp in length
* 22966 24501: contig of 1436 bp in length

24502 24601: gap of unknown length
 * 24602 26428: contig of 1827 bp in length
 * 26429 26528: gap of unknown length
 * 26529 28503: contig of 1975 bp in length
 * 28504 28603: gap of unknown length
 * 28604 30124: contig of 1521 bp in length
 * 30125 30224: gap of unknown length
 * 30225 32292: contig of 2068 bp in length
 * 32293 32393: gap of unknown length
 * 32393 34644: contig of 2252 bp in length
 * 34645 34744: gap of unknown length
 * 34745 37728: contig of 2984 bp in length
 * 37729 37828: gap of unknown length
 * 37829 40230: contig of 2402 bp in length
 * 40231 40330: gap of unknown length
 * 40331 42294: contig of 1964 bp in length
 * 42295 42394: gap of unknown length
 * 42395 44875: gap of 2481 bp in length
 * 44876 44975: contig of unknown length
 * 44976 48629: contig of 3654 bp in length
 * 48630 48729: gap of unknown length
 * 48730 51512: contig of 2783 bp in length
 * 51513 51612: gap of unknown length
 * 51613 54010: contig of 2398 bp in length
 * 54011 54110: gap of unknown length
 * 54111 55956: contig of 1846 bp in length
 * 55957 56056: gap of unknown length
 * 56057 57509: contig of 1453 bp in length
 * 57510 57610: gap of unknown length
 * 57610 63737: gap of 6128 bp in length
 * 63738 63837: gap of unknown length
 * 63838 66881: contig of 3044 bp in length
 * 66882 70918: gap of unknown length
 * 66982 70918: contig of 3937 bp in length
 * 70919 75925: gap of unknown length
 * 75926 76025: contig of 4907 bp in length
 * 76026 76940: gap of unknown length
 * 76941 79041: contig of 2915 bp in length
 * 79041 84797: gap of unknown length
 * 84798 84898: contig of 5757 bp in length
 * 84898 88965: gap of unknown length
 * 88966 89065: contig of 4068 bp in length
 * 89066 92420: gap of unknown length
 * 92421 92520: gap of 3355 bp in length
 * 92521 97435: gap of unknown length
 * 97436 102886: contig of 4815 bp in length
 * 97436 102886: contig of 5451 bp in length
 * 102887 102986: gap of unknown length
 * 102987 106824: contig of 3838 bp in length
 * 106825 106924: gap of unknown length
 * 106925 114199: contig of 7275 bp in length
 * 114200 114299: gap of unknown length
 * 114300 120990: contig of 6691 bp in length
 * 120991 121090: gap of unknown length
 * 121091 129531: contig of 8441 bp in length
 * 129532 129631: gap of unknown length
 * 129632 134993: contig of 5262 bp in length
 * 134993 134993: gap of unknown length
 * 134994 142772: contig of 7779 bp in length
 * 142773 142872: gap of unknown length
 * 142873 157121: contig of 14249 bp in length
 * 157122 157221: gap of unknown length
 * 157222 164701: contig of 7480 bp in length
 * 164702 164801: gap of unknown length
 * 164802 171481: contig of 6680 bp in length
 * 171482 171581: gap of unknown length
 * 171582 184008: contig of 12427 bp in length
 * 184009 184108: gap of unknown length
 * 184109 192392: contig of 8284 bp in length
 * 192393 192492: gap of unknown length
 * 192493 203020: contig of 10528 bp in length
 * 203021 203120: gap of unknown length

203121 211871: contig of 8751 bp in length
 * 211872 211971: gap of unknown length
 * 211972 223920: contig of 11949 bp in length
 * 223921 224020: gap of unknown length
 * 224021 237820: contig of 13800 bp in length
 * 237821 237920: gap of unknown length
 * 237921 256627: contig of 18707 bp in length
 * 256628 256727: gap of unknown length
 * 256728 280655: contig of 23928 bp in length
 * 280656 280755: gap of unknown length
 * 280756 304407: contig of 23652 bp in length.
 Location/Qualifiers
 1. .304407
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-6981"
 /clone_id="RP23 mouse BAC library 23"

ORIGIN
 Query Match 1.2% Score 22; DB 2; Length 304407;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCTCAGCACACGCTGTCAGG 457
 |||||
 Db 5763 GCTCAGCACACGCTGTCAGG 5784

RESULT 15
 AEO05048
 LOCUS
 Halobacterium sp. NRC-1 section 79 of 170 of the complete genome.
 DEFINITION
 AEO05048 AEO04437
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Halobacterium sp. NRC-1
 Halobacterium sp. NRC-1
 Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
 Halobacteriaceae; Halobacterium.
 1 (bases 1 to 10225)
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
 Shukla,H.D., Lasky,S.R., Baliga,N., Thorson,V., Sbrogna,J.,
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
 Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angeline,C.M., Dale,H.,
 Ienbacher,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
 Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
 Dassa,S.
 Genome sequence of Halobacterium species NRC-1
 Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
 MEDLINE
 PUBMED
 20504483
 11016950

REFERENCE
 AUTHORS
 2 (bases 1 to 10225)
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
 Shukla,H.D., Lasky,S.R., Baliga,N., Thorson,V., Sbrogna,J.,
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
 Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angeline,C.M., Dale,H.,
 Ienbacher,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
 Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
 Dassa,S.
 Direct Submission
 Submitted (14-JUL-2000) Institute for Systems Biology, 4225
 Roosevelt Way NE, Seattle, WA 98105, USA
 Location/Qualifiers
 1. .10225
 /organism="Halobacterium sp. NRC-1"
 /mol_type="genomic DNA"
 /strain="NRC-1"
 /db_xref="taxon:64091"

FEATURES
 source

gene 125. .1510
/gene="huth"
/note="synonym: VNG1212G"
CDS 125. .1510
/gene="huth"
/note="huth"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase"
/protein_id="AAG19580.1"
/db_xref="GI:10580744"
/translation="MSDTRIDAADREAIQANIVRSHAAGSELDTAAVRALLVTELN
ALAKYSGIRERVDLVGLNAGVPPVPSRSGASGDLAIAMSRVLGEQAD
VGERMPAAEALAAADLEPTLQAKGALINGTOLTGYAALALVDAEVRASADTA
GALTTEVMTSTAACA PAIHEVRPHQGNVSAHIRTNLITAGSEVLDDHDCRVDAY
SIRCLPQVIGAVARDALDHLRAVATELNSNDPVPYFAGGRVSPRASGTDRAVSGG
NRFGEVLARLGTYASALAEALAISSRTDRLLNPETQEPHLEPFIAPDSGLHSLMT
PQYTAASLVNDLSLQPTLDNASVSGAQEDHVSMSGAAYNFRBAVKAATVGVGL
LCGAQGREFLDPDLALAGTAAAYDLVREVESEVAGDRALADMAAVGDLVRAGVED
AVARALDAFLA"
gene 1598. .2578
/gene="VNG1213C"
CDS 1598. .2578
/gene="VNG1213C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1213c"
/protein_id="AAG19581.1"
/db_xref="GI:10580745"
/translation="MSAAQOTVYRSACRSGGRASDIGNPLPTATFDRLLCHCATARR
PTSECAPDTERASAPRTREPETALPMCOPOAVDMRVNSFLPAAGVGERTERLMQGV
THMTDGDGPGIGDATANIYSGFIDAEARALDAGNTSPFADLPSELRLRYENAVD
DVAFPDIEETTGDDHSSVYTVVSVHKGQDTTIVRGSUDLAERLQALLDAPMYANG
KRPDAFLERHFDLSLPHLDLMPCKRLDLTGAKQVERDLGVRAEDVDGREA
VRLMRRYASDNDDEAALDLVLYNQDTQNLQALLDAVBERLHRDVEPHV"
gene 3236. .4366
/gene="VNG1214H"
CDS 3236. .4366
/gene="VNG1214H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1214h"
/protein_id="AAG19582.1"
/db_xref="GI:10580746"
/translation="MTSHDYRLAEIEDGPERLAIDLTLTRDRYGVDPDQGRGKDG
GKVDLLDGDGDNIIIVHVSRRBDMQKLTDLGKAADHDYDIFVYVYNR.IITNQ
KPVDAOPFVDEYGEIDIWDERLSELDNNHQDLREERYLIADEDDPTGKTALI
DEBLRLRRRDNELPRVNDAPYAVLHVPVAVDASGDTDRHDLPTPLPGRYGYGS
YKNTLDGRVLYAAGDNEPRGYIYVDTGGMVAVADAPMGNEENMTIGQSPFKLIGD
AIEYGNALNELELDGPFVEVGLSVLSVKGYSPFTKKGCGGDRNGPVRFRNDIEAKPH
TVEDTDSPTGEAKRGFDRVWGRWSDSPYSDNEWHER"
gene 4638. .4638
/gene="trn19"
tRNA complement (4655. .4638)
/gene="trn19"
/product="tRNA-Thr"
complement (4650. .5132)
/gene="pal1"
/note="synonym: VNG1215G"
complement (4650. .5132)
/gene="pal1"
/note="pal1"
/codon_start=1
/transl_table=11
/product="sporulation regulator homolog"
/protein_id="AAG19583.1"
/db_xref="GI:10580747"
/translation="MTVRPASGTDIDAVVAELMWQLAEQRAHSGHLLAENRAQARDL
IGQYVADGVAVTQAGRTVGFWFHTETGFIYDTRGTIDLIVAPGRGGRGSA
LTHAERELRSRGADALAIETALDNNAAARLYERRGYAPHRVTFEQSVEVDTHSKDDP
"

gene complement (5129. .6325)
/gene="pgk"
/note="synonym: VNG1216G"
CDS complement (5129. .6325)
/gene="pgk"
/note="pgk"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate isomerase"
/protein_id="AAG19584.1"
/db_xref="GI:10580748"
/translation="MAIRLIDLDAANRAIGVRVDINSPLTAAGGLADARLRRAVDT
LAEILDAARVAVLAAHGRPGGDEBARLERHADRLDALDAEVSYCDATFSGARAV
ALAPAEAVLNTETFESEVYAPAEARADTALVDGLAPALDAYVNDAPFAAHRGAP
SLVGPEVPEVPSAGRVMEALDLSGVADEPTPTPTTYVGGAARVPSVEVAALASHGL
ANVNLVTGVAVNPLAATGVDDIGRSTDEIHERDGTETARAADLLAANNDAHLHPVD
VAVRGAACEISTDLPPAGAEAVCDISDVTDAADVADLADETVVVGPRGVEFDD
LRADRGVGFDAASEVESHIVGGGDTAAIRRFDTITGFHVSTGGGAINTLTDADLP
AVAAALR"
gene complement (6428. .7144)
/gene="VNG1218C"
CDS complement (6428. .7144)
/gene="VNG1218C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1218c"
/protein_id="AAG19585.1"
/db_xref="GI:10580749"
/translation="MQRFAGCPTAVMAVALGPPRPSVHGFFYRGLLYRRMNDVRLKE
ALSASRVTVSARAADVAATTLREHATTAVQORAGDLGVTADAVDAVADGHLSD
STVADVMAVDTEVSLVDATVADADAMARTGDGCVLVVDADAELAGVSAOAILALA
ANMDRGRGSRPLGPDPRPVGASGAPDPPEPSNGDPTTAGANTYSSQVCERCG
LAAALGRVNRGLCTECSRSL"
gene 7107. .7853
/gene="urk"
/note="synonym: VNG1219G"
CDS 7107. .7853
/gene="urk"
/note="urk"
/codon_start=1
/transl_table=11
/product="uridine kinase"
/protein_id="AAG19586.1"
/db_xref="GI:10580750"
/translation="MTAVGVQAKRCIQHASPPRAVSMITPSPALIGAGTGAKTIV
ABEITDNVESATLLPLDNYEDLSDRPEERANAYDPSAFEWELRTHIDALSG
OSIEMEQVDPERHAKADRVVPEPDVIVLEGIALLSDETVNDMLDLHIVETDADVR
ILRLREVDVERHRELEGVMDYLVSTKMHQGLEPTGRHADIIIPEGANSVAVML
EKVQAESEMTAMARDDTAFEDATVYNDQ"
gene complement (7911. .8234)
/gene="VNG1220H"
CDS complement (7911. .8234)
/gene="VNG1220H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1220h"
/protein_id="AAG19587.1"
/db_xref="GI:10580751"
/translation="MAEMEDPSGRDVMHDPDGEHSGEGGRKGLAVLSKMTVEGEDFP
LDAAEFVAIEGYEYVRLNQTIVSVADIPEHIDADBEFTKVGPFHKTUVGAALDDADPLD
YHPDA"
gene complement (8533. .8615)
/gene="trn20"
tRNA complement (8533. .8615)
/gene="trn20"
/product="tRNA-Ser"
8926. .10119
/gene="orc8"
/note="synonym: VNG1224G"
8926. .10119
/gene="orc8"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 16:25:39 ; Search time 479.978 seconds
(without alignments)
16621.841 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878
Sequence: 1 atcaggtctcagggtctgtcc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N.GeneSeq_29Jan04.*

1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1878	100.0	1878	4	AAFP25374
2	963	51.3	3275	4	AAFP25373
3	24	1.3	27	4	AAFP25372
4	20	1.1	23	4	AAFP25371
5	20	1.1	596	4	AAFP12773
6	20	1.1	700	4	AAH93017
7	20	1.1	47475	3	AAH81465
8	20	1.1	110000	3	AAH81489_5
9	20	1.1	154902	6	ABQ88198
10	20	1.1	349980	3	AAFP21612
11	19	1.0	481	3	AAH68016
12	19	1.0	481	9	ADDA1766
13	19	1.0	484	4	AAK64246
14	19	1.0	485	2	AAV23847
15	19	1.0	495	2	AAZ06848
16	19	1.0	495	3	AAH69597
17	19	1.0	495	3	AAH67934
18	19	1.0	495	3	ADDA1684
19	19	1.0	501	6	AAH61626
20	19	1.0	538	6	ABT10660
21	19	1.0	1341	6	ABZ14660
22	19	1.0	1838	4	ABL11339
23	19	1.0	1876	4	ABL15067

24	19	1.0	2000	2	AAV33671	AAV33671 Candida k
25	19	1.0	2000	2	AAV33672	AAV33672 Candida k
26	19	1.0	2000	4	AAH73590	AAH73590 Candida k
27	19	1.0	2000	4	AAH73589	AAH73589 Candida k
28	19	1.0	2000	7	ACA62356	ACA62356 Gene enco
29	19	1.0	2000	7	ACA62357	ACA62357 mRNA enco
30	19	1.0	2000	8	ACA62409	ACA62409 C. krusei
31	19	1.0	2000	8	ACA62410	ACA62410 C. krusei
32	19	1.0	2195	4	AAK87051	AAK87051 Human imm
33	19	1.0	2204	4	AB126576	AB126576 Drosophila
34	19	1.0	2462	4	AAK87052	AAK87052 Human imm
35	19	1.0	2535	4	AAK53267	AAK53267 Human pol
36	19	1.0	2565	4	AAK52283	AAK52283 Human pol
37	19	1.0	2674	4	AB115134	AB115134 Drosophila
38	19	1.0	2832	6	ABN79893	ABN79893 Fungal ZB
39	19	1.0	3579	7	ABT11898	ABT11898 Human bre
40	19	1.0	4172	4	AB111397	AB111397 Drosophila
41	19	1.0	4357	4	AB111398	AB111398 Drosophila
42	19	1.0	4365	4	AB115066	AB115066 Drosophila
43	19	1.0	4387	4	AB113810	AB113810 Drosophila
44	19	1.0	5742	4	AB103491	AB103491 Drosophila
45	19	1.0	6804	4	AB113824	AB113824 Drosophila

ALIGNMENTS

RESULT 1	AAFP25374	standard; cDNA; 1878 BP.
ID	AAFP25374	
XX	AAFP25374	
AC	AAFP25374	
XX	15-MAY-2001	(first entry)
DT		
XX		
DE		Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
XX		
KM		SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW		viral resistance; resistance; fatty acid content; protein content; ss.
XX		
OS		Arabidopsis thaliana.
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1878
FT		/*tag= a
FT		/product= "SGS3"
XX		
PN	WO200105951-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	13-JUL-2000; 2000MO-FR002052.	
XX		
PR	16-JUL-1999; 99FR-00009417.	
PR	26-JAN-2000; 2000FR-00001006.	
XX		
PA	(AVET) AVENTIS CROPS SCIENCE SA.	
PA	(INRG) INST NAT RECH AGRONOMIQUE.	
XX		
PI	Beclin C, Elmayan T, Vaucheret H;	
XX		
DR	WPI: 2001-159529/16.	
XX	P-FSDB; AAB11798.	
PT		
PT	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus	
PT	resistance in plants and, when inhibited, for increasing transgene	
XX	expression.	
XX		
PS	Claim 1; Page 32-35; 36pp; French.	
XX		
CC	The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.	
CC	The SGS3 gene is essential for post-transcriptional inactivation	
CC	(degradation of RNA) and for resistance to viruses. Overexpression of	

XX	AF25373;
XX	AC
XX	15-MAY-2001 (first entry)
XX	Genomic sequence of the Arabidopsis SGS3 gene.
XX	DE
XX	SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX	KM
XX	viral resistance; resistance; fatty acid content; protein content; ss.
XX	OS
XX	Arabidopsis thaliana.
XX	PH
XX	Key
XX	FT
XX	exon
XX	Location/Qualifiers
XX	696..1658
XX	/tag= a
XX	FT
XX	intron
XX	1659..1731
XX	/tag= b
XX	FT
XX	exon
XX	1732..2023
XX	/tag= c
XX	FT
XX	intron
XX	2024..2134
XX	/tag= d
XX	FT
XX	exon
XX	2135..2379
XX	/tag= e
XX	FT
XX	intron
XX	2380..2481
XX	/tag= f
XX	FT
XX	exon
XX	2482..2648
XX	/tag= g
XX	FT
XX	intron
XX	2649..2738
XX	/tag= h
XX	FT
XX	exon
XX	2739..2949
XX	/tag= i
XX	FT
XX	WO200105951-A2.
XX	PN
XX	25-JAN-2001.
XX	PD
XX	13-JUL-2000; 2000MO-FR002052.
XX	PE
XX	16-JUL-1999; 99FR-00009417.
XX	PR
XX	26-JAN-2000; 2000FR-00001006.
XX	XX
XX	PA
XX	(AVET) AVENTIS CROPS SCIENCE SA.
XX	(INRG) INST NAT RECH AGRONOMIQUE.
XX	PA
XX	Beclin C, Elmayan T, Vaucheret H;
XX	PI
XX	DR
XX	WPI; 2001-159529/16.
XX	XX
XX	FT
XX	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX	FT
XX	resistance in plants and, when inhibited, for increasing transgene
XX	FT
XX	expression.
XX	XX
XX	PS
XX	Claim 1; Page 31-32; 36pp; French.
XX	XX
XX	The present sequence represents the genomic sequence of the Arabidopsis
XX	CC
XX	thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
XX	CC
XX	inactivation (degradation of RNA) and for resistance to viruses.
XX	CC
XX	Overexpression of SGS3 results in plants with increased resistance to
XX	CC
XX	viruses, while inactivation of SGS3 in transgenic plants (e.g. by
XX	CC
XX	expressing antisense RNA, by mutation or by homologous recombination)
XX	CC
XX	increases the level of the transgene product. This product may e.g.
XX	CC
XX	impart resistance (to herbicide, insects or pathogens), alter contents of
XX	CC
XX	essential fatty acids or proteins, or is pharmaceutically active, e.g. an
XX	CC
XX	immunoglobulin or interferon
XX	SO
XX	Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
XX	Query Match
XX	Best Local Similarity 51.3%; Score 963; DB 4; Length 3275;
XX	Matches 963; Conservative 100.0%; Pred. No. 0;
XX	Mismatches 0; Indels 0; Gaps 0
XX	1 ATGAGTTCTAGGGCTGCTCCATGCTCTAGAGAAAGAAAGCTTCAGGCTGTTATAGGCT 60

Db	696	ATGAGTCTTAGGGGCTGGTCCAAATGCTTAAGGAAAAAAGACGTTACAGGGTGGTTAATGACCT	755
QY	61	GAGGTGGAACAGTTGGTTCCAAAGGTTTGGCAGAGGACGAGACTGGCTTCTTCAAGAGATGAT	120
Db	756	GAGGTGGAACAGTTGGTTCCAAAGGTTTGGCAGAGGACGAGACTGGCTTCTTCAAGAGATGAT	815
QY	121	GGAGGAGAGTGGGAGGTCAATTTCCAAAGAAAGAACAAACCAAGGAAACACTTCTGGA	180
Db	816	GGAGGAGAGTGGGAGGTCAATTTCCAAAGAAAGAACAAACCAAGGAAACACTTCTGGA	875
QY	181	AAAACCTTGGGTTTCTCGAATTCGAAATTCCTCTAGACTTGGGGTGTGACAGCAAGGG	240
Db	876	AAAACCTTGGGTTTCTCGAATTCGAAATTCCTCTAGACTTGGGGTGTGACAGCAAGGG	935
QY	241	AGAGGTACCAACGTAATCTGGAGAGGAAACAAATGATTCGGGAGAGGTAAACGGCAATGCT	300
Db	936	AGAGGTAGCAACGTAATCTGGAGAGGAAACAAATGATTCGGGAGAGGTAAACGGCAATGCT	995
QY	301	CGGGGCAATTCAGCTAACATATCTGTCGGGGAGAGACCGTTGAGCAGAAAAGTAAGATTAAC	360
Db	996	CGGGGCAATTCAGCTAACATATCTGTCGGGGAGAGACCGTTGAGCAGAAAAGTAAGATTAAC	1055
QY	361	AACTTTGGGCAACCCCAACCTGTATCTCGCCCTCTTGGAAAGAGATGGAATTTGGCAG	420
Db	1056	AACTTTGGGCAACCCCAACCTGTATCTCGCCCTCTTGGAAAGAGATGGAATTTGGCAG	1115
QY	421	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGAGAGAGTTTCTGACGTGGAGAGATGAT	480
Db	1116	GCAAGAGAGAGTTCTGCTCAGACACACAGCTGTGAGAGAGTTTCTGACGTGGAGAGATGAT	1175
QY	481	GTGATTAATGCTTCTGAGGAAAGAAATGATTCGATCCTTTGGATGATTCGATGAGAC	540
Db	1176	GTGATTAATGCTTCTGAGGAAAGAAATGATTCGATCCTTTGGATGATTCGATGAGAC	1235
QY	541	CTTGCAGATGATGATTAATGACTCGGATGTGAGTCAAAAAGAGCCATGATCAAGAAACGAG	600
Db	1236	CTTGCAGATGATGATTAATGACTCGGATGTGAGTCAAAAAGAGCCATGATCAAGAAACGAG	1295
QY	601	AATAAGTGTTCMAAAAAGTTCTTTGGCAGCTTGGAGTAGCTTGATGCTGATCGACAGATAAT	660
Db	1296	AATAAGTGTTCMAAAAAGTTCTTTGGCAGCTTGGAGTAGCTTGATGCTGATCGACAGATAAT	1355
QY	661	GAACCAAGAGGCGAGTGGCATTTGTCAGCTTGTCAAGACCGAAGCTGGTGCATGATTTGG	720
Db	1356	GAACCAAGAGGCGAGTGGCATTTGTCAGCTTGTCAAGACCGAAGCTGGTGCATGATTTGG	1415
QY	721	TATTAACCTGCACCCCTCTACTAGCTCATGCGAGCAAAAAGGAGCTAGGCGAGTTAAGCTC	780
Db	1416	TATTAACCTGCACCCCTCTACTAGCTCATGCGAGCAAAAAGGAGCTAGGCGAGTTAAGCTC	1475
QY	781	CATAGAGAAATGGCTGGAAGTTTGAAGAAAGGATCTACAGATGAGAGGCGCATCTGTCAAT	840
Db	1476	CATAGAGAAATGGCTGGAAGTTTGAAGAAAGGATCTACAGATGAGAGGCGCATCTGTCAAT	1535
QY	841	CCTTGTGTGATTTATATGCGCAGTGAAGGCTTTGGGTGAGATGA AAAAGATTTATGA	900
Db	1536	CCTTGTGTGATTTATATGCGCAGTGAAGGCTTTGGGTGAGATGA AAAAGATTTATGA	1595
QY	901	ATTGTGCGGCTCCAAATGTCATCATATGAATCTAGACTGGAATTAAGGACGATTAAGAT	960
Db	1596	ATTGTGCGGCTCCAAATGTCATCATATGAATCTAGACTGGAATTAAGGACGATTAAGAT	1655
QY	961	AAG 963	
Db	1656	AAG 1658	
<p>RESULT 3 AAF25372/c ID AAF25372 standard; DNA: 27 BP. AC XX AAF25372; XX</p>			

DT 15-MAY-2001 (first entry)
 XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
 DE
 XX
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;
 KM viral resistance; resistance; fatty acid content; protein content;
 KM PCR primer; ss.
 OS Arabidopsis thaliana.
 XX WO200105951-A2.
 PN 25-JAN-2001.
 XX
 PD 13-JUL-2000; 2000WO-FR002052.
 XX
 PF 16-JUL-1999; 99FR-00009417.
 XX 26-JAN-2000; 2000FR-00001006.
 PR
 XX (AVET) AVENTIS CROPS SCIENCE SA.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 PI Beclin C, Elmayer T, Vaucheret H;
 XX WPI; 2001-159529/16.
 DR
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 PT
 PS Example 1; Page 22; 36pp; French.
 XX
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
 CC transcriptional inactivation (degradation of RNA) and for resistance to
 CC viruses. Overexpression of SGS3 results in plants with increased
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants
 CC (e.g. by expressing antisense RNA, by mutation or by homologous
 CC recombination) increases the level of the transgene product. This product
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter
 CC contents of essential fatty acids or proteins, or is pharmaceutically
 CC active, e.g. an immunoglobulin or interferon
 CC
 XX Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
 SQ
 Query Match 1.3%; Score 24; DB 4; Length 27;
 Best Local Similarity 100.0%; Pred. No. 0.71;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1855 GGCCCTTCACATGAGATGATGTA 1878
 DB 27 GGCCCTTCACATGAGATGATGTA 4

RESULT 4
 AAF25371
 ID AAF25371 standard; DNA; 23 BP.
 XX
 AC AAF25371;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
 XX
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;
 KM viral resistance; resistance; fatty acid content; protein content;
 KM PCR primer; ss.
 OS Arabidopsis thaliana.
 XX WO200105951-A2.
 PN 25-JAN-2001.
 XX
 PD

XX
 PF 13-JUL-2000; 2000WO-FR002052.
 XX
 PR 16-JUL-1999; 99FR-00009417.
 XX 26-JAN-2000; 2000FR-00001006.
 PR
 XX (AVET) AVENTIS CROPS SCIENCE SA.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 PI Beclin C, Elmayer T, Vaucheret H;
 XX WPI; 2001-159529/16.
 DR
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 PT
 PS Example 1; Page 22; 36pp; French.
 XX
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
 CC transcriptional inactivation (degradation of RNA) and for resistance to
 CC viruses. Overexpression of SGS3 results in plants with increased
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants
 CC (e.g. by expressing antisense RNA, by mutation or by homologous
 CC recombination) increases the level of the transgene product. This product
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter
 CC contents of essential fatty acids or proteins, or is pharmaceutically
 CC active, e.g. an immunoglobulin or interferon
 CC
 XX Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;
 SQ
 Query Match 1.1%; Score 20; DB 4; Length 23;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAGTTCTAGGCTGCTCC 20
 DB 4 ATGAGTTCTAGGCTGCTCC 23

RESULT 5
 AAF12773
 ID AAF12773 standard; cDNA; 596 BP.
 XX
 AC AAF12773;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:5296.
 XX
 XX Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US007781.
 XX
 PR 22-MAR-1999; 99US-00273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 DR

XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX
 PS Claim 88, Page 2210; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered.
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination.
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organisation of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 CC *Trichoderma reesei*, which are all specifically claimed in the present
 CC invention
 XX
 SQ Sequence 596 BP; 144 A; 168 C; 150 G; 134 T; 0 U; 0 Other;
 XX
 Query Match 1.1%; Score 20; DB 3; Length 596;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 142 TCCAGAGAGAGAGAGAGAA 161
 Db 431 TCCAGAGAGAGAGAGAGAA 450
 XX
 RESULT 6
 ID AAH93017/c
 AC AAH93017, standard; DNA; 700 BP.
 XX
 AC AAH93017,
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Human inflammatory bowel disease related gene fragment IGR3310a.
 XX
 KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
 KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
 XX chromosome 5q31-33; forensic test; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200142511-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 11-DEC-2000; 2000WO-US033632.
 XX
 PR 10-DEC-1999; 99US-0170257P.
 PR 10-APR-2000; 2000US-0196046P.
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (ELLT-) ELLIPISIS BIOTHERAPEUTICS CORP.
 XX
 PI Daly M, Hudson TJ, Lander ES, Rioux J, Simionovitch K;
 XX

DR WPI; 2001-367874/38.
 XX
 PT Testing for the presence of polymorphisms associated with inflammatory
 PT bowel disease, using a hybridization assay.
 XX
 PS Disclosure, Page 381; 463pp; English.
 XX
 CC The present invention describes a method for detecting the presence of
 CC polymorphisms associated with inflammatory bowel diseases such as
 CC ulcerative colitis and Crohn's disease. The methods can be used to detect
 CC the presence of genetic polymorphisms associated with inflammatory bowel
 CC disease and correlating their occurrence with disease states. They may be
 CC used in this way for phenotypic correlations, forensics, paternity
 CC testing, medicine and genetic analysis. The present sequence is a gene
 CC containing a polymorphic site described in the exemplification of the
 CC invention
 XX
 SQ Sequence 700 BP; 156 A; 158 C; 149 G; 237 T; 0 U; 0 Other;
 XX
 Query Match 1.1%; Score 20; DB 4; Length 700;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1696 TTTCAGAGAGAGAGATGGA 1715
 Db 415 TTTCAGAGAGAGAGATGGA 396
 XX
 RESULT 7
 ID AAA81465/c
 AC AAA81465, standard; DNA; 47475 BP.
 XX
 AC AAA81465;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_13 SEQ ID NO:13.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US023573.
 XX
 PR 09-OCT-1998; 98US-0103794P.
 PR 30-APR-1999; 99US-0132068P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI; 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
 PT in the diagnosis and treatment of *N. meningitidis* infection and other
 PT Neisserial infections, for example, *N. gonorrhoea*.
 XX
 PS Claim 7; Page 383-397; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic proteins
 CC from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent
 CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;
 CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA
 CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
 CC AAA81304 to AAA81321 represent PCR primers used in the isolation of

CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent
 CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
 CC used in the exemplification of the present invention. The nucleic acid
 CC sequences, protein sequences, and antibodies against them, can be used in
 CC the manufacture of a composition. The composition can be used as a
 CC medicament (or in the manufacture of a medicament) for treating,
 CC preventing or diagnosing infection due to Neisserial bacteria. For
 CC example, some of the identified proteins could be components of vaccines
 CC against Meningococcus B; against all serotypes; and/or against all
 CC pathogenic Neisseriae. Identification of sequences from the bacterium
 CC will also facilitate production of biological probes, particularly
 CC organism-specific probes. Attempts to make efficacious Meningococcus B
 CC vaccines have failed mainly due to antigen tolerance. Multivalent
 CC vaccines have also been tried but none have successfully overcome
 CC antigenic variability. The provision of further, complete sequences may
 CC provide an opportunity to identify secreted or surface exposed proteins
 CC that may be presumed targets for the immune system and which are not
 CC antigenically variable or at least more conserved than other more
 CC variable regions

XX SQ Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;
 XX Query Match 1.1%; Score 20; DB 3; Length 47475;
 XX Best Local Similarity 100.0%; Pred. No. 55;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGTTCAAGTTT 86
 DB 23354 GAACAGTTGTTCAAGTTT 23335

RESULT 8
 AAA81489_5/c of AAA81489 from base 500001 (N. meningitidis partial DNA sequence
 Continuation (6 of 9) of fragments LOCUS AAA81489 Accession Aaa81489
 WP Sequence split into 9 fragments Begin End
 WP Fragment Name 1 110000
 WP AAA81489_1 100001 210000
 WP AAA81489_2 200001 310000
 WP AAA81489_3 300001 410000
 WP AAA81489_4 400001 510000
 WP AAA81489_5 500001 610000
 WP AAA81489_6 600001 710000
 WP AAA81489_7 700001 810000
 WP AAA81489_8 800001 837096

Query Match 1.1%; Score 20; DB 3; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGTTCAAGTTT 86
 DB 32595 GAACAGTTGTTCAAGTTT 32576

RESULT 9
 AB088198/c
 ID AB088198 standard; cDNA; 154902 BP.
 AC AB088198;
 XX 18-SEP-2002 (first entry)
 DE Human osteoblast differentiation related cDNA SEQ ID NO 105.
 XX Human; osteoblast; stem cell differentiation; bone tissue deposition;
 KM osteoporosis; osteopathic; ss.
 OS Homo sapiens.
 XX WO200250301-A2.
 XX 27-JUN-2002.
 PD

XX PF 18-DEC-2001; 2001WO-US048276.
 XX PR 18-DEC-2000; 2000US-0255882P.
 XX PR 24-APR-2001; 2001US-0285691P.
 XX (GENE-) GENE LOGIC INC.
 XX (PROC) PROCTER & GAMBLE CO.
 XX PI D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
 XX Mertz L;
 XX WPI; 2002-557663/59.
 XX Use of genes and their expression profiles associated with osteoblast
 XX differentiation for screening modulators bone formation, for diagnosing
 XX or treating e.g. osteoporosis, or as markers for the differentiation
 XX process.

XX Claim 1; SEQ ID NO 105; 78bp + Sequence Listing; English.

XX The invention relates to genes and their expression profiles are used
 XX for: (a) screening modulators of precursor stem cell differentiation into
 XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
 XX deposition of bone tissue, abnormal rate of osteoblast formation or
 XX osteoporosis; or (c) treating or monitoring of bone tissue deposition;
 XX cited in (b), or monitoring the progression of bone tissue deposition;
 XX specific conditions include postmenopausal osteoporosis, glucocorticoid
 XX osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
 XX induced abnormalities in bone formation or bone loss, conditions that
 XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
 XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
 XX or fibrous dysplasia. The present sequence is that of an osteoblast
 XX differentiation associated cDNA marker of the invention. Note: The
 XX sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;
 XX Query Match 1.1%; Score 20; DB 6; Length 154902;
 XX Best Local Similarity 100.0%; Pred. No. 54;
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1712 TGGAGAGTTGTGGAAGG 1731
 DB 128153 TGGAGAGTTGTGGAAGG 128134

RESULT 10
 AAF21612/c
 ID AAF21612 standard; DNA; 349980 BP.
 XX AAF21612;
 XX 13-MAR-2001 (first entry)
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.
 OS Neisseria meningitidis.
 XX WO200066791-A1.
 XX 09-NOV-2000.
 XX 08-MAR-2000; 2000WO-US005928.
 XX 30-APR-1999; 99US-0132068P.
 XX 08-OCT-1999; 99WO-US023573.
 PR

```

PR 28-FEB-2000; 2000GH-00004695.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Piza M, Hickey E, Peterson J, Tetteijn H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Frazer CM, Grandi G;
XX
XX WPI; 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections.
XX
XX
XX Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of Neisseria
CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
CC represent fragments of the NMB genomic sequence, as the sequence was too
CC long to go in a record on its own it was split into 8 sequences which
CC overlap each other at the beginning and end of each sequence by 49980 bp
CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
CC given in AAB588550 to AAB588593, and AAF21589 to AAF21606 represent PCR
CC primers which are used in the exemplification of the present invention.
CC The NMB genome and fragments from it have antibacterial activity, and can
CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
CC and/or antibodies which binds to the proteins can be used in compositions
CC for treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computer, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
CC outer membrane proteins currently used
XX
XX Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;
SQ
XX
XX Query Match 1.1%; Score 20; DB 3; Length 349980;
XX Best Local Similarity 100.0%; Pred. No. 54;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
XX 67 GAACAGTTGGTTCAAGGTTT 86
Db 167824 GAACAGTTGGTTCAAGGTTT 167805
XX
XX
XX RESULT 11
XX AAA68016/c
XX AAA68016 standard; DNA; 481 BP.
XX
XX AAA68016;
XX
XX 24-OCT-2000 (first entry)
XX
XX Eucalyptus grandis CCR nucleotide sequence SEQ ID NO:109.
XX
XX Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;
XX Pinus radiata; Monterey pine; ds.
XX
XX Eucalyptus grandis.
XX
XX WO200022099-A1.
XX
XX 20-APR-2000.
XX
XX 06-OCT-1999; 99WO-NZ000168.
XX
XX 09-OCT-1998; 98US-00169789.
XX 14-JUL-1999; 99US-0143811P.
XX
XX

```

PA	(GENE-) GENESIS RES & DEV CORP LTD.
PB	FLETCHER CHALLENGE FORESTS LTD.
XX	
XX	Bloksberg LN, Havukkala IU;
XX	
DR	WP1; 2000-317962/27.
XX	
PT	Novel polynucleotide encoding enzymes involved in lignin-biosynthetic pathway useful for producing transgenic plants especially eucalyptus and pine species having altered lignin content, composition and structure.
PT	
XX	
PS	Claim 1; Page 91-92; 213pp; English.
CC	
CC	The present invention describes isolated polynucleotides and proteins encoding and representing the enzymes cinnamate 4-hydroxylase (C4H), coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase (OMT), cinammyl alcohol dehydrogenase (CAD), cinamoyl-CoA reductase (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4Cl), coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG), laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase, caffeic acid methyl transferase, caffeine CoA methyl transferase, coumarate CoA ligase, cytochrome P450 LXX1A, diphenol oxidase, flavanol glucosyl transferase, flavenoid hydroxylase, and isoflavone reductase, which are involved in the lignin biosynthetic pathway. The polynucleotides can be used for modulating lignin content, lignin composition and the structure of a plant, especially eucalyptus and pine species, and for modifying the activity of an enzyme involved in lignin biosynthetic pathway, and for producing a plant having altered lignin content, composition and structure. They can be used for designing probes and primers useful for detecting similar DNA and RNA sequences in any organism and for PCR amplification. The lignin content can be efficiently modified using the polynucleotides. AAB67908 to AAB68201 and AAB16341 to AAB16449 represent polynucleotide and protein sequences used in the exemplification of the present invention
SO	
XX	Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;
QY	
DB	Query Match 1.0%; Score 19; DB 3; Length 481; Best local similarity 100.0%; Pred. No. 1.8e+02; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
YY	74 TTGGTTCAAGGTTTGCCAGG 92 470 TTGGTTCAAGGTTTGCCAGG 452
RESULT 12	
ID	ADD41766 standard; DNA; 481 BP.
XX	
AC	ADD41766;
XX	
DT	15-JAN-2004 (first entry)
XX	
DE	Cinnamoyl-CoA reductase DNA #21.
XX	
KW	des; plant; lignin content; lignin composition; lignin structure; lignin biosynthesis; gene.
XX	
OS	Eucalyptus grandis.
XX	
FN	US2003131373-A1.
PD	10-JUL-2003.
XX	
PF	18-JUN-2002; 2002US-00174693.
XX	
PR	11-SEP-1996; 96US-00713000. 21-NOV-1997; 97US-00975316. 09-OCT-1998; 98US-00169789. 14-JUL-1999; 99US-0143833P. 12-JUL-2000; 2000US-00615192.
PR	
XX	

PA (BLOCK/) BLOKSEBERG L N.
PA (HAYU/) HAYUKKALA I.
XX
PI Bloksberg LN, Hayukkala I;
XX
XX WPI; 2003-829606/77.
DR
XX
XX New polynucleotide associated with the lignin biosynthetic pathway, for
PT modulating lignin content, composition and structure of plants, or
PT producing a plant with altered lignin content, composition and structure.
PI
XX
XX Claim 1; SEQ ID NO 109; 2066P; English.
PS
XX
XX The invention relates to an isolated polynucleotide. The polynucleotides,
CC polypeptides and genetic constructs are useful for modulating lignin
CC content, composition and structure of plants, or for producing a plant
CC having altered lignin content, composition and structure. The
CC polynucleotides are also useful in genome mapping, physical mapping,
CC positional cloning of genes, or as non-disruptive tags for marking
CC organisms, particularly plants. The present sequence represents a
CC polynucleotide associated with the lignin biosynthetic pathway.
CC
XX
SQ Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;

Query Match 1.0%; Score 19; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAGGTTGGCAGG 92
DB 470 TGGTTCAGGTTGGCAGG 452
|||||

RESULT 13
AAK64246
ID AAK64246 standard; CDNA; 484 BP.
XX
AC AAK64246;
XX
DT 06-NOV-2001 (first entry).
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:9306.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metacelais; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0226279P.
PR 18-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-02277009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229345P.
PR 01-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX P-PSDB; AAM91465.
 PT
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating cancers and metastasis.
 PS Claim 1; SEQ ID NO 9306; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 484 BP; 179 A; 66 C; 106 G; 131 T; 0 U; 2 Other;
 Query Match 1.0%; Score 19; DB 4; Length 484;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1767 GGAAGACATGAGAGAGG 1785
 DB 323 GGAAGACATGAGAGAGG 341
 RESULT 14
 AAV23847/C
 ID AAV23847 standard; DNA; 495 BP.
 XX
 AC AAV23847;
 XX
 DT 31-JUL-1998 (first entry)
 XX
 DE Plant CCR enzyme DNA sequence.
 XX
 XX Lignin biosynthetic pathway; eucalyptus; pine; transgenic plant;
 KW Lignin content; tree processing; cellulose fibre; ss.
 XX
 OS Eucalyptus grandis.
 XX
 XX WO9811205-A2.
 XX
 XX 19-MAR-1998.
 XX
 PF 10-SEP-1997; 97WO-NZ000112.
 XX
 PR 11-SEP-1996; 96US-00713000.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Blokberg LN, Grierson AW, Havukkala IU;
 XX
 XX WPI; 1998-207374/18.
 DR
 XX
 PT Sequences useful for modification of plant lignin content or structure -
 PT from Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) are
 PT associated with lignin biosynthesis pathway, useful e.g. in paper
 PT industry.
 PS Claim 1; Page 39; 82pp; English.
 XX
 CC This sequence represents a fragment of the CCR enzyme coding sequence. It
 CC is an example of a DNA sequence of the invention, which are from
 CC Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) associated with
 CC the lignin biosynthesis pathway. Constructs containing the DNA sequences
 CC can be used to produce transgenic plants or plant cells, especially woody
 CC plants e.g. eucalyptus or pine species but also e.g. monocotyledons or
 CC dicotyledons; by stably incorporating the constructs into the plant
 CC genome. The lignin content or structure, or activity of a specific enzyme
 CC in the plant, can therefore be modulated. Reductions in lignin content or
 CC changes in composition are useful in tree processing for paper. High
 CC lignin content results in energy- and chemical-intensive separation
 CC methods in order to obtain the pure cellulose fibre required. Reductions
 CC in lignin content may also be useful for forage crops, whilst increases
 CC or changes in composition may be desirable to increase the mechanical
 CC strength of wood, change its colour or increase its resistance to rot.
 CC The sequences are also useful as probes to isolate DNA sequences encoding
 CC enzymes involved in the lignin biosynthesis pathway from other plant
 CC species
 XX
 SQ Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;
 Query Match 1.0%; Score 19; DB 2; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTTCAAGTTTGGCAGG 92
 |||||
 DB 484 TGGTTCAAGTTTGGCAGG 466

Search completed: April 9, 2004, 02:09:06
 Job time : 482.978 secs

RESULT 15
 AAZ06848/C
 ID AAZ06848 standard; cDNA, 495 BP.

XX AAZ06848;
 AC
 XX

DT 09-NOV-1999 (first entry)

XX Eucalyptus cinamoyl-CoA reductase (CCR) partial cDNA 2.

XX Lignin; biosynthesis; forage crop; wood; paper production;
 KM transgenic plant; ss.

XX Eucalyptus grandis.

XX US5952486-A.

XX 14-SEP-1999.

XX 21-NOV-1997; 97US-00975316.

XX 11-SEP-1996; 96US-00713000.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Grierson AW, Bloksberg LN, Havukkala I;

XX WPI; 1999-527029/44.

XX Isolated DNA sequence encoding enzymes from the lignin synthetic pathway
 useful for generating plants with an altered lignin content.

XX Example 1; Col 37-38; 48pp; English.

XX This sequence represents a cinamoyl-CoA reductase (CCR) partial cDNA
 CC from Eucalyptus grandis. This enzyme is involved in the biosynthesis of
 CC lignin, an insoluble polymer which is primarily responsible for the
 CC rigidity of plant stems. Lignin serves as a matrix around the
 CC polysaccharide components of some plant cell walls. The higher the lignin
 CC content, the more rigid the plant. Lignin also plays a role in disease
 CC resistance of plants by impeding the penetration and propagation of
 CC pathogenic agents. Lignin is formed by polymerisation of at least three
 CC different monolignols (para-coumaryl alcohol, coniferyl alcohol and
 CC sinapyl alcohol). These three monolignols are synthesised by similar
 CC pathways from phenylalanine in a multistep process and are believed to be
 CC polymerised into lignin via a free radical mechanism. The lignin content
 CC of plants can be altered using DNA sequences encoding these enzymes.
 CC Lignin content can be increased by incorporation of additional copies of
 CC genes encoding these enzymes into the target plant. This could be
 CC beneficial for increasing the mechanical strength of wood. Similarly, a
 CC decrease in lignin content can be obtained by transforming the target
 CC plant with antisense copies of such genes. This may be beneficial in
 CC plants used as forage crops for livestock (lignin is indigestible) and in
 CC trees used in paper manufacture

XX Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;

XX Query Match 1.0%; Score 19; DB 2; Length 495;

XX Best local Similarity 100.0%; Pred. No. 1.8e+02;

XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTTCAAGTTTGGCAGG 92
 |||||

DB 484 TGGTTCAAGTTTGGCAGG 466

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 20:08:06 ; Search time 3159.03 Seconds
(without alignments)
1752.632 Million cell updates/sec

Title: US-10-030-829-2
Perfect score: 1878
Sequence: 1 atgagttctcagggctgtctc.....ttccaatgaagatgatga 1878

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
EST:
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	565	30.1	693	28	AQ959658 LEREM21TF
2	534	28.4	606	28	AQ959657 LEREM21TF
3	513	27.3	619	9	AV566465 AV566465
4	512	27.3	647	9	AU236368 AU236368

Result No.	Score	Query Match	Length	DB ID	Description
5	474	25.2	613	28	AQ964580 LERGX20TF
6	329	17.5	380	29	AV525508 AV525508
7	295	15.7	457	29	CNS00029H Arabidops
8	273	14.5	523	28	AQ959659 LEREM21TF
9	244	13.0	421	9	AU227277 AU227277
10	228	12.1	290	28	AQ010650 AU227277
11	169	9.0	548	9	AI999551 AI999551
12	148	8.1	197	28	BH169457 SALK 0013
13	153	7.9	443	28	BH618393 SALK 0013
14	148	7.9	443	28	BH618393 SALK 0013
15	148	7.9	443	28	BH618393 SALK 0013
16	148	7.9	443	28	BH618393 SALK 0013
17	148	7.9	443	28	BH618393 SALK 0013
18	148	7.9	443	28	BH618393 SALK 0013
19	148	7.9	443	28	BH618393 SALK 0013
20	148	7.9	443	28	BH618393 SALK 0013
21	148	7.9	443	28	BH618393 SALK 0013
22	148	7.9	443	28	BH618393 SALK 0013
23	148	7.9	443	28	BH618393 SALK 0013
24	148	7.9	443	28	BH618393 SALK 0013
25	148	7.9	443	28	BH618393 SALK 0013
26	148	7.9	443	28	BH618393 SALK 0013
27	148	7.9	443	28	BH618393 SALK 0013
28	148	7.9	443	28	BH618393 SALK 0013
29	148	7.9	443	28	BH618393 SALK 0013
30	148	7.9	443	28	BH618393 SALK 0013
31	148	7.9	443	28	BH618393 SALK 0013
32	148	7.9	443	28	BH618393 SALK 0013
33	148	7.9	443	28	BH618393 SALK 0013
34	148	7.9	443	28	BH618393 SALK 0013
35	148	7.9	443	28	BH618393 SALK 0013
36	148	7.9	443	28	BH618393 SALK 0013
37	148	7.9	443	28	BH618393 SALK 0013
38	148	7.9	443	28	BH618393 SALK 0013
39	148	7.9	443	28	BH618393 SALK 0013
40	148	7.9	443	28	BH618393 SALK 0013
41	148	7.9	443	28	BH618393 SALK 0013
42	148	7.9	443	28	BH618393 SALK 0013
43	148	7.9	443	28	BH618393 SALK 0013
44	148	7.9	443	28	BH618393 SALK 0013
45	148	7.9	443	28	BH618393 SALK 0013

ALIGNMENTS

RESULT 1
AQ959658/c
LOCUS
DEFINITION
AQ959658 LERE Arabidopsis thaliana genomic clone LEREM21, genomic survey sequence.
ACCESSION
AQ959658
VERSION
AQ959658.1 GI:6787359
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@igir.org
For additional information, see <http://www.tigr.org/tdb/at.html>

Seq primer: TF
 Class: Shotgun.
 Location/Qualifiers
 1.693

/organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERECTA"
 /db_xref="taxon:3702"
 /clone_lib="LEREM21"
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was
 sheared to 0.6-0.8 Kbp before ligation."

ORIGIN

Query Match 30.1%; Score 565; DB 28; Length 693;
 Best Local Similarity 99.8%; Pred. No. 1.6e-276; Mismatches 1; Indels 0; Gaps 0;
 Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

134 AGGTCAATTCAGAGAAAGAAACAAACCAAGAAACATCTTCTGAAAAAATTGGGTTT 193
 |||||
 688 AGGTCAATTCAGAGAAAGAAACAAACCAAGAAACATCTTCTGAAAAAATTGGGTTT 629
 |||||
 194 CTCAGAAATTCGAATCCTCTTGAAGCTTGGGGTGTCAAGCAAGGAGAGAGAGCAAG 253
 |||||
 628 CTCAGAAATTCGAATCCTCTTGAAGCTTGGGGTGTCAAGCAAGGAGAGAGAGCAAG 569
 |||||
 254 TATCTGGAGAGAGAAACATGTATCCGGAGAGGTAACGGCAATGTCGGGCAATTCAG 313
 |||||
 568 TATCTGGAGAGAGAAACATGTATCCGGAGAGGTAACGGCAATGTCGGGCAATTCAG 509
 |||||
 314 CTAAATATCTGTGTGGGAGAGAGCTTGAAGCAAAAGTATGATTAACATCTTGTGGC 373
 |||||
 508 CTAAATATCTGTGTGGGAGAGAGCTTGAAGCAAAAGTATGATTAACATCTTGTGGC 449
 |||||
 374 CCCCACTGTATCTGGGCTCTTGGAGAGAGATGAATTTGGCAGGCAAGAGAGTT 433
 |||||
 448 CCCCACTGTATCTGGGCTCTTGGAGAGAGATGAATTTGGCAGGCAAGAGAGTT 389
 |||||
 434 CTGCTCAGACACAGAGCTGTGAGAGATTTCTGAAGTGTGAGATGATGTGATTAATCTT 493
 |||||
 388 CTGCTCAGACACAGAGCTGTGAGAGATTTCTGAAGTGTGAGATGATGTGATTAATCTT 329
 |||||
 494 CTGAGAGAGAGATGATTCGATCTTGTGATGATTCGATGAGAGACCTTGCAGATGATG 553
 |||||
 328 CTGAGAGAGAGATGATTCGATCTTGTGATGATTCGATGAGAGACCTTGCAGATGATG 269
 |||||
 554 ATTATGATCTGGAGTGTGATCAAAAGAGCATGATCAAGAAACAGATTAAGTGTCA 613
 |||||
 268 ATTATGATCTGGAGTGTGATCAAAAGAGCATGATCAAGAAACAGATTAAGTGTCA 209
 |||||
 614 AAAAGTTCTTGGAGCTTGGATGATCTTGTGATGAGAGATTAATGAACCAAGAGGC 673
 |||||
 208 AAAAGTTCTTGGAGCTTGGATGATCTTGTGATGAGAGATTAATGAACCAAGAGGC 149
 |||||
 674 AGTGCATTTGTCCACTTGTCAAGAGAGACCTGAGCATGATGATTAATCACTGCACC 733
 |||||
 148 AGTGCATTTGTCCACTTGTCAAGAGAGACCTGAGCATGATGATTAATCACTGCACC 89
 |||||
 734 CTCTACTAGCTCATGC 749
 |||||
 88 CTCTACTAGCTCATGC 73

RESULT 2
 A0959657/ 606 bp DNA linear GSS 28-JAN-2000
 LOCUS LEREM21TF LERE Arabidopsis thaliana genomic clone LEREM21, genomic
 DEFINITION survey sequence.
 ACCESSION A0959657
 VERSION A0959657.1 GI:6787358
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS
 TITLE
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T.,
 Feldblyum, T., Liang, F., Creasy, T., and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: atc@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Seq primer: TF

FEATURES
 source
 Location/Qualifiers
 1.606
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERECTA"
 /db_xref="taxon:3702"
 /clone_lib="LEREM21"
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was
 sheared to 0.6-0.8 Kbp before ligation."

ORIGIN

Query Match 28.4%; Score 534; DB 28; Length 606;
 Best Local Similarity 100.0%; Pred. No. 1e-260; Mismatches 0; Indels 0; Gaps 0;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

216 AGCTTGGGGTGTGACGACGAGAGAGAGTNGCAACCTATCTGGAGAGAGAAACAATGT 275
 |||||
 606 AGCTTGGGGTGTGACGACGAGAGAGAGTNGCAACCTATCTGGAGAGAGAAACAATGT 547
 |||||
 276 ATCCGGAGAGAGTAAACGCAATGTGCGGGCACTTCAACTAATATCTGTCGGGAGC 335
 |||||
 546 ATCCGGAGAGAGTAAACGCAATGTGCGGGCACTTCAACTAATATCTGTCGGGAGC 487
 |||||
 336 AGCGTTGACAGAAAGTATGATTAACAATCTTGTGCAACCCCACTGTATCTCGCCCTCC 395
 |||||
 486 AGCGTTGACAGAAAGTATGATTAACAATCTTGTGCAACCCCACTGTATCTCGCCCTCC 427
 |||||
 396 TTTGAGAGAGATGGAATTTGGCAGCAAGAGAGGTTCTGCTCAGACACAGCTGTGCA 455
 |||||
 426 TTTGAGAGAGATGGAATTTGGCAGCAAGAGAGGTTCTGCTCAGACACAGCTGTGCA 367
 |||||
 456 GGAATTTCTGACGCTGAGAGATGATGATTAATGCTTCTGAGAGAGAGATGATTCGCA 515
 |||||
 366 GGAATTTCTGACGCTGAGAGATGATGATTAATGCTTCTGAGAGAGAGATGATTCGCA 307
 |||||
 516 TGCTTGTGATGATTTGTGATGACGACCTTGCAGATGATTAATGATCTCGAGATGATGA 575
 |||||
 306 TGCTTGTGATGATTTGTGATGACGACCTTGCAGATGATTAATGATCTCGAGATGATGA 247
 |||||
 576 AAAGAGCATGATGATCAAGAGCAATTAAGTGTCAAAAAGTTCTTTGGAGCTTGA 635
 |||||
 246 AAAGAGCATGATGATCAAGAGCAATTAAGTGTCAAAAAGTTCTTTGGAGCTTGA 187
 |||||
 636 TAGCTTGTGATGATGATTAATGAACCAAGAGAGAGCTGATTTGCTCAGCTTGTCA 695
 |||||
 186 TAGCTTGTGATGATGATTAATGAACCAAGAGAGAGCTGATTTGCTCAGCTTGTCA 127
 |||||
 696 GAACGACCTGTGTGCTCATGATTTGATTAACCTGCACCCCTCTACTAGCTCATGC 749
 |||||
 126 GAACGACCTGTGTGCTCATGATTTGATTAACCTGCACCCCTCTACTAGCTCATGC 73

RESULT 3

AV566465/c 619 bp mRNA linear EST 07-SEP-2000
 LOCUS AV566465 Arabidopsis thaliana green siliques Columbia Arabidopsis
 DEFINITION thaliana cDNA clone SQ244b06f 3', mRNA sequence.
 ACCESSION AV566465
 VERSION AV566465.1 GI:8737917
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 619)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7 (3), 175-180 (2000)
 JOURNAL MEDLINE
 PUBMED 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 location/Qualifiers
 1..619
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ244b06F"
 /tissue_type="green siliques"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /note="Vector: BluescriptII SK-, Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN

Query Match 27.3%; Score 513; DB 9; Length 619;
 Best Local Similarity 100.0%; Pred. No. 5.5e-250;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1291 CAAGAGTGGTTGTAAGAGCTGAGCAGATCTGTGAGCAATCAGACGTGAATAC 1350
 619 CAGAGATGGTTGTAAGAGCTGAGCAGATCTGTGAGCAATCAGACGTGAATAC 560
 1351 TTTAAGAACAGCTCTCAAAACAGAACAGACCCCAAGGTCCTTGAGAAATCTTGAA 1410
 559 TTTAAGAACAGCTCTCAAAACAGAACAGACCCCAAGGTCCTTGAGAAATCTTGAA 500
 1411 ATTATAGCGAAGAGCTGCTTAAGCTGAGAGATTAATGCGATCTGAGAACAGAACT 1470
 499 ATTATAGCGAAGAGCTGCTTAAGCTGAGAGATTAATGCGATCTGAGAACAGAACT 440
 1471 AAGATCAGCATGAACAGAACAGAGAGATGATGATCAGACAGAGTTTTCATGAT 1530
 439 AAGATCAGCATGAACAGAACAGAGAGATGATGATCAGACAGAGTTTTCATGAT 380
 1531 TCATCAACAGATTCATGAAGAGAGAGCGCAAGAGAGAAATTTGAGATGTTGAG 1590
 379 TCATCAACAGATTCATGAAGAGAGAGCGCAAGAGAGAAATTTGAGATGTTGAG 320
 1591 CACGAGAGAGCGTCCAAAGTGTGTGGCAGACAGAGAGAACTTAATCCCTCTAGAAAT 1650
 319 CACGAGAGAGCGTCCAAAGTGTGTGGCAGACAGAGAGAACTTAATCCCTCTAGAAAT 260
 1651 GAGCATTCGGAAGAGAGCTGAGAGATGTCAGATTCATGAGTTTCAAGAGAAAGAG 1710
 259 GAGCATTCGGAAGAGAGCTGAGAGATGTCAGATTCATGAGTTTCAAGAGAAAGAG 200
 1711 ATGAGAGAGTTTGTGAGAGAGAGAGATCTGATTAAGATCAAGAGAGAGATGAGAA 1770
 199 ATGAGAGAGTTTGTGAGAGAGAGAGATCTGATTAAGATCAAGAGAGAGATGAGAA 140

QY 1771 GACATGAGAGAGAGCATCAGAGAGATTTT 1803
 DB 139 GACATGAGAGAGAGCATCAGAGAGATTTT 107

RESULT 4
 LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002
 DEFINITION AU236368 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-93-K05 5',
 mRNA sequence.
 ACCESSION AU236368
 VERSION AU236368.1 GI:19875537
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 647)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
 and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished (2002)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msek@rcc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified lambda FLIC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
 details.

ORIGIN

Query Match 27.3%; Score 512; DB 9; Length 647;
 Best Local Similarity 100.0%; Pred. No. 1.8e-249;
 Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTTCAAGCTGCTGCTCAATGCTTAAGAGAAAGAAAGAGTTCAGAGTGTATAGGCT 60
 92 ATGAGTTTCAAGCTGCTGCTCAATGCTTAAGAGAAAGAAAGAGTTCAGAGTGTATAGGCT 151
 61 GAGGTGAACAGTGTGCTTCAAGTTTGAGAGGAGCAGACTGGCTTCTCAAGATGAT 120
 92 ATGAGTTTCAAGCTGCTGCTCAATGCTTAAGAGAAAGAAAGAGTTCAGAGTGTATAGGCT 151
 152 GAGGTGAACAGTGTGCTTCAAGTTTGAGAGGAGCAGACTGGCTTCTCAAGATGAT 211
 121 GAGAGAGAGTGGAGGTCATTTTCAAGAGAAACAGAACAAACAGAGAAACACTTGTGA 180
 212 GAGAGAGAGTGGAGGTCATTTTCAAGAGAAACAGAACAAACAGAGAAACACTTGTGA 271
 181 AAAACCTGGTTTTCAGAAATCGAATCCTCTAGAGCTTGGGAGTGTGAGAGCAAGGG 240
 272 AAAACCTGGTTTTCAGAAATCGAATCCTCTAGAGCTTGGGAGTGTGAGAGCAAGGG 331
 241 AGAGTAGCAAGTATCTGAGAGAGAGAAACAATGTATCCGAGAGAGTAAACGCAATGCT 300

Db 332 AGAGGTAGCAAGTATCTGGAGAGAAACAATGATCCGAGAGGTACGGCAATGGT 391
 QY 301 CGGGGCAATCAAGCTAACATATCTGTGGGGAGAGGCTTGAGCAGAAATATGATAC 360
 Db 392 CGGGGCAATCAAGCTAACATATCTGTGGGGAGAGGCTTGAGCAGAAATATGATAC 451
 QY 361 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGGAATTTGGAG 420
 Db 452 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGGAATTTGGAG 511
 QY 421 GCAAGAGAGGTTCTGTCTGAGCACAAGCTGAGCAGAGATTCTGACGTGGAGGATAT 480
 Db 512 GCAAGAGAGGTTCTGTCTGAGCACAAGCTGAGCAGAGATTCTGACGTGGAGGATAT 571
 QY 481 GTGATTAATGCTTCTGAGAGAGAAATGATTC 512
 Db 572 GTGATTAATGCTTCTGAGAGAGAAATGATTC 603

RESULT 5
 LOCUS AQ964580/c 613 bp DNA linear GSS 28-JUN-2000
 DEFINITION LERGX20TF LERG Arabidopsis thaliana genomic clone LERGX20, genomic
 survey sequence.
 ACCESSION AQ964580
 VERSION AQ964580.1 GI:6792281

KEYWORDS Arabidopsis thaliana (chale crees)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 613)
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utechtach, T.,
 Feldblum, T., Liang, F., Creasy, T., and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoyang Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/cdb/at/cdb.html>
 Seq primer: TF
 Class: shotgun.

FEATURES

source location/Qualifiers
 1..613
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERECTA"
 /db_xref="taxon:3702"
 /clone="LERGX20"
 /note="Organ: Leaf; Vector: pUC19UK, Total genomic DNA was
 sheared to 0.4-0.7 kbp before ligation."

ORIGIN

Query Match 25.2%; Score 474; DB 28; Length 613;
 Best Local Similarity 99.8%; Pred. No. 4.5e-220; Indels 0; Gaps 0;
 Matches 524; Conservative 1;
 1 ATGAGTTAGAGGCTGCTCAATGCTTAAGAAAAAAGAGCTTCAAGGCTGTTATAGGCTT 60
 Db 525 ATGAGTTAGAGGCTGCTCAATGCTTAAGAAAAAAGAGCTTCAAGGCTGTTATAGGCTT 466
 QY 61 GAGGTTGAACAGTGTGTTCAAGTTGGCAGAGAGAGAGCTGCTTCTTCAACAAGATGAT 120
 Db 465 GAGGTTGAACAGTGTGTTCAAGTTGGCAGAGAGAGAGCTGCTTCTTCAACAAGATGAT 406

QY 121 GGAGGAGAGTGAGAGGTCATTTCCAGAGAAACAAGAAACAACAGAAACACTTCTGA 180
 Db 405 GGAGGAGAGTGAGAGGTCATTTCCAGAGAAACAAGAAACAACAGAAACAACACTTCTGA 346
 QY 181 AAAAATCTGGGTTTTCAGAAATTCGAAATCTCTTAAGCTTGGGTTGTCAGCAGCAAGG 240
 Db 345 AAAAATCTGGGTTTTCAGAAATTCGAAATCTCTTAAGCTTGGGTTGTCAGCAGCAAGG 286
 QY 241 AGAGGTAGCAACATATCTGAGAGAGAAACAATGATCCGGAGAGGTTAACCGCAATGCT 300
 Db 285 AGAGGTAGCAACATATCTGAGAGAGAAACAATGATCCGGAGAGGTTAACCGCAATGCT 226
 QY 301 CGGGGCAATCAAGCTAACATATCTGTGGGGAGCAGCGTTGACGAAAGATATGATTAAC 360
 Db 225 CGGGGCAATCAAGCTAACATATCTGTGGGGAGCAGCGTTGACGAAAGATATGATTAAC 166
 QY 361 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGGAATTTGGAG 420
 Db 165 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGGAATTTGGAG 106
 QY 421 GCAAGAGAGGTTCTGTCTGAGCACAAGCTGAGCAGAGATTCTGACGTGGAGGATAT 480
 Db 105 GCAAGAGAGGTTCTGTCTGAGCACAAGCTGAGCAGAGATTCTGACGTGGAGGATAT 46
 QY 481 GTGATTAATGCTTCTGAGAGAGAAATGATTCGATGCTTTGGAT 525
 Db 45 GTGATTAATGCTTCTGAGAGAGAAATGATTCGATGCTTTGGAT 1

RESULT 6
 LOCUS AV525508 380 bp mRNA linear EST 01-SEP-2000
 DEFINITION AV525508 Arabidopsis thaliana aboveground organs two to six-week
 old Arabidopsis thaliana cDNA clone APD25d02R 5', mRNA sequence.

ACCESSION AV525508
 VERSION AV525508.1 GI:8685036
 KEYWORDS EST
 SOURCE Arabidopsis thaliana (chale crees)
 ORGANISM Arabidopsis thaliana
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 380)
 Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7 (3), 175-180 (2000)

JOURNAL MEDLINE 20363093
 PUBMED 10907847
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>

FEATURES

source location/Qualifiers
 1..380
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="APD25d02R"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /clone_lib="Arabidopsis thaliana aboveground organs two to
 six-week old"
 /note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN

Query Match 17.5%; Score 329; DB 9; Length 380;
 Best Local Similarity 99.7%; Pred. No. 4e-156;
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 276 ATCCGGAGAGGTAAACGGCAATGTCGGGCAATTCAGCTAATATCTGTCGGGACG 335
 DB 1 ATCCGGAGAGGTAAACGGCAATGTCGGGCAATTCAGCTAATATCTGTCGGGACG 60
 QY 336 AGGCTTGAGCAGAAAGTATATTAACAATTGTCGACCCCACTGTATCTGCTCC 395
 DB 61 AGGCTTGAGCAGAAAGTATATTAACAATTGTCGACCCCACTGTATCTGCTCC 120
 QY 396 TTGGAAGAGAGATGAAATTTGCAAGGAGGAGGTTCTGCTCAGCAGCTGTCGA 455
 DB 121 TTGGAAGAGAGATGAAATTTGCAAGGAGGAGGTTCTGCTCAGCAGCTGTCGA 180
 QY 456 GGAGTTTCTGACGTGAGAGATGATGATTAATGCTTGAAGAGAGATATTCGA 515
 DB 181 GGAGTTTCTGACGTGAGAGATGATGATTAATGCTTGAAGAGAGATATTCGA 240
 QY 516 TGCCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGTGTGACTCA 575
 DB 241 TGCCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGTGTGACTCA 300
 QY 576 AAGAGCCATGATTCAGAAAGCAAGATAGTGTGTTCAAAAAGTTCTTGGCAGCTTGA 635
 DB 301 AAGAGCCATGATTCAGAAAGCAAGATAGTGTGTTCAAAAAGTTCTTGGCAGCTTGA 360
 QY 636 TAGCTTGTGATGACGACGAGA 655
 DB 361 TAGCTTGTGATGACGACGAGA 380

RESULT 7
 CENS00P9H 457 bp DNA linear GSS 28-JUN-1999
 LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC F8G21 of
 DEFINITION IGF library from strain Columbia of Arabidopsis thaliana, genomic
 survey sequence.

ACCESSION AL084227.1 GI:5285367
 VERSION AL084227
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 457)
 AUTHORS Salanoubat, M., Choisme, N., Artiguenave, F., Brottier, P., Wincker, P.,
 Samson, D., Saurin, W., Weissenbach, J., and Quetier, F.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 457)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

FEATURES
 source 1..457
 Location/Qualifiers
 1..457
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_11b="F8G21"
 /clone_11b="IGF"
 /note="end : T7"

ORIGIN

Query Match 15.7%; Score 295; DB 29; Length 457;
 Best Local Similarity 100.0%; Pred. No. 9.8e-139;
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGGCTCGGCAAGGCAACAGAGCTGCTGAATCTTGCACAGATAGAGGCTTTA 1021
 DB 55 AGTGGCTCGGCAAGGCAACAGAGCTGCTGAATCTTGCACAGATAGAGGCTTTA 114

QY 1022 GAGCAGCCATTCCTATGCTCCAGAGGCCATCTGAGATGACTGTTGATGTTGAGA 1081
 DB 115 GAGCAGCCATTCCTATGCTCCAGAGGCCATCTGAGATGACTGTTGATGTTGAGA 174
 QY 1082 GCAGTCCACTGGCTATTTGAGAGCCGAACGCTTCCACCGGAGTTAGCTGAGATGGGT 1141
 DB 175 GCAGTCCACTGGCTATTTGAGAGCCGAACGCTTCCACCGGAGTTAGCTGAGATGGGT 234
 QY 1142 TAGATGAATGCTGGGCTCAGAAAGCAGATGATGTTTCTGAGAGGTTGGCAACTGT 1201
 DB 235 TAGATGAATGCTGGGCTCAGAAAGCAGATGATGTTTCTGAGAGGTTGGCAACTGT 294
 QY 1202 ATGCTTCTTGCAAGCAAGACAGATCTGACATATTCATCAACACTCAAG 1256
 DB 295 ATGCTTCTTGCAAGCAAGACAGATCTGACATATTCATCAACACTCAAG 349

RESULT 8
 A0959659 523 bp DNA linear GSS 28-JAN-2000
 LOCUS LEREM21R LERE Arabidopsis thaliana genomic clone LEREM21, genomic
 DEFINITION survey sequence.

ACCESSION A0959659
 VERSION A0959659.1 GI:6787360
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 523)
 AUTHORS Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T.,
 Feldlyum, T., Liang, F., Creasy, T., and Fraser, C.M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atc@igr.org
 For additional information, see <http://www.tigr.org/cdb/at.html>
 Seq primer: TR
 Class: shotgun.

FEATURES
 source 1..523
 Location/Qualifiers
 1..523
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERRECTA"
 /db_xref="taxon:3702"
 /clone="LEREM21"
 /clone_11b="LERE"

ORIGIN
 Query Match 14.5%; Score 273; DB 28; Length 523;
 Best Local Similarity 99.3%; Pred. No. 1.8e-127;
 Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 sheared to 0.6-0.8 kbp before ligation."

QY 167 GAAACACTTCTGAAAACCTTGGGTTTCAGAAATTCGAATCTCTTAAGCTTTGGGTG 226
 DB 25 GAAACACTTCTGAAAACCTTGGGTTTCAGAAATTCGAATCTCTTAAGCTTTGGGTG 84

QY 227 GTCAGCAGCAAGGAGAGGTAGCAACGTATCTGGAGAGAAACAATGTATCCGAGAG 286
 DB 85 GTCAGCAGCAAGGAGAGGTAGCAACGTATCTGGAGAGAAACAATGTATCCGAGAG 144

QY 287 GTAACGGCAATGCTGGGCAATTCAGGTAACATATCTGCTGGGAGCAGCGTTGAGCA 346
 DB 145 GTAACGGCAATGCTGGGCAATTCAGGTAACATATCTGCTGGGAGCAGCGTTGAGCA 204

QY 347 GAAAGTATGATTAACAATTGTGGACACCCCACTGTATCTGCCCTCTTGGAGAG 406
DB 205 GAAAGTATGATTAACAATTGTGGACACCCCACTGTATCTGCCCTCTTGGAGAG 264
QY 407 GATGAATTGGCAGGCAAGAGAGTTCTGCTCAGACACAGCTGTGAGAGTTCCG 466
DB 265 GATGAATTGGCAGGCAAGAGAGTTCTGCTCAGACACAGAGTTGCGAGAGTTCTG 324
QY 467 ACCTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 526
DB 325 ACCTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384
QY 527 ATTCTGATGACGACCTTGGACAGTATGATGATGATGATGATGATGATGATGATG 586
DB 385 ATTCTGATGACGACCTTGGACAGTATGATGATGATGATGATGATGATGATGATG 444
QY 587 GATCAC 592
DB 445 GATCAC 450

RESULT 9
AU227277/c 421 bp mRNA linear EST 23-APR-2002
LOCUS AU227277 RAF14 Arabidopsis thaliana cDNA clone RAF14-93-K05 3'
DEFINITION mRNA sequence.
ACCESSION AU227277 GI:19741924
VERSION AU227277
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 421)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A.,
Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rct.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLX-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified Bluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source location/Qualifiers
1..421
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF14-93-K05"
/issue_type="root"
/lab_host="DH10B"
/clone_1lb="RAF14"
/note="Site_1: BamHI; Site_2: SalI"

ORIGIN
Query Match 13.0%; Score 244; DB 9; Length 421;
Best Local Similarity 100.0%; Pred.No.1e-112;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1635 TAATCCTTAGCAATGACGATTGCCGAAAGAGAGCTGAGGAAGTGTCAAGCTTATCGA 1694

DB 421 TAATCCTTAGCAATGACGATTGCCGAAAGAGAGCTGAGGAAGTGTCAAGCTTATCGA 362
QY 1695 GTTTCAGAGAAAGATGAGAGAGTTGTGGAAGAGAGAGATGCTGATAAAGATCA 1754
DB 361 GTTTCAGAGAAAGATGAGAGAGTTGTGGAAGAGAGAGATGCTGATAAAGATCA 302
QY 1755 AGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAG 1814
DB 301 AGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAG 242
QY 1815 AGAATTTGATGAGCTTTGGAACAGCTCATGTCACAGCATGAGCTTCAAGTAAGATGA 1874
DB 241 AGAATTTGATGAGCTTTGGAACAGCTCATGTCACAGCATGAGCTTCAAGTAAGATGA 182
QY 1875 TTGA 1878
DB 181 TTGA 178

RESULT 10
AQ010650 290 bp DNA linear GSS 29-MAY-1998
LOCUS F27C8TRC IGF Arabidopsis thaliana genomic clone F27C8, genomic
DEFINITION survey sequence.
ACCESSION AQ010650
VERSION AQ010650
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 290)
AUTHORS Rounsley,S.D., Sub.E.J., Wible,C., Golden,K., Shatsman,S., Choi,P.,
Yu,K., Akintoye,B., Shen,K., Goonasekaram,S., Millscher,J.,
Adams,M.D. and Venter,J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in
JOURNAL Arabidopsis Genomic Sequencing. Update 4
COMMENT Other GSSs: F27C8TRC
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@igr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 290.
Location/Qualifiers
1..290
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F27C8"
/sex="hermaphrodite"
/clone_1lb="IGF"
/note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 12.1%; Score 228; DB 28; Length 290;
Best Local Similarity 100.0%; Pred.No.1.3e-104;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 962 AGTGCTCGCATGCGGCAACCAAGAGCTGTGGAATTAATTGACAGATGAGGCTCTTA 1021
DB 63 AGTGCTCGCATGCGGCAACCAAGAGCTGTGGAATTAATTGACAGATGAGGCTCTTA 122
QY 1022 GAGACGCGCATTTCTTATGTCACAGGAGCATGCTGAGATGAGTGTTCGATGTTGAGA 1081

Db 123 GACACCCCATCTCTAATGCTCCACAGGCCATCGTGGAGATGATGTTCTGATGTTTGAGA 182

Qy 1082 GCAGTGCATCGCTATTTGAGGCGGAGCCCTCCACGGGAGTTAGCTGAGATGGGGT 1141

Db 183 GCAGTGCATCGCTATTTGAGGCGGAGCCCTCCACGGGAGTTAGCTGAGATGGGGT 242

Qy 1142 TAGATGAATTCCTGCGGCTCAGAGCGCAGTATGTTTCTGAGAGTG 1189

Db 243 TAGATGAATTCCTGCGGCTCAGAGCGCAGTATGTTTCTGAGAGTG 290

RESULT 11
AI999551/c
LOCUS 548 bp mRNA linear EST 08-SEP-1999
DEFINITION thaliana cDNA clone 701556368, mRNA sequence.
ACCESSION AI999551
VERSION AI999551.1 GI:5846456
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 548)
AUTHORS Chen, J., Montiyama, M., Chan, E., Mooney, M., Carron, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mounauou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kasutny, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobriga, A., Murry, L., Turner, C., Kikorian, S., Elder, L. and Hanson, D.
TITLE Arabidopsis thaliana Gene Expression Microarray
JOURNAL Unpublished (1999)
COMMENT Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
LOCATION/Qualifiers
1. 548
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia Col-0"
/db_xref="taxon:3702"
/clone="701556368"
/issue_type="rosette"
/dev_stage="4 - 7 weeks"
/clone_lib="A. thaliana, Columbia Col-0, rosette-3"
/note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunt-ended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

ORIGIN
Query Match 9.0%; Score 169; DB 9; Length 548;
Best Local Similarity 100.0%; Pred. No. 2.2e-74;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1710 GATGAGGAGCTTTGTGGAAGAGGAGGATGCTGATAAAGATCAAGAGAAGATGGA 1769

Db 345 GATGAGGAGCTTTGTGGAAGAGGAGGATGCTGATAAAGATCAAGAGAAGATGGA 286

Qy 1770 AGACATGAAGAAGAGCATCAAGAGAGATATTGATCTGAGAAAGATTTGATGAGGC 1829

Db 285 AGACATGAAGAAGAGCATCAAGAGAGATATTGATCTGAGAAAGATTTGATGAGGC 226

Qy 1830 TTGGAACAGCTCAGTACAGATGCGCTTCAACAATGAATGATGA 1878

Db 225 TTGGAACAGCTCAGTACAGATGCGCTTCAACAATGAATGATGA 177

RESULT 12
BH169457
LOCUS 197 bp DNA linear GSS 03-OCT-2001
DEFINITION SALK_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_001394, genomic survey sequence.
ACCESSION BH169457
VERSION BH169457.1 GI:15904832
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 197)
AUTHORS Alonso, J.M., Leisbe, T.J., Barajas, P., Chen, H., Cheuk, R., Gadriab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shim, P., Zimmerman, J. and Ecker, D.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
THIS is single pass sequence recovered from the left border of TDNA.
Class: TDNA tagged.
LOCATION/Qualifiers
1. 197
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_001394"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
Query Match 8.1%; Score 153; DB 28; Length 197;
Best Local Similarity 100.0%; Pred. No. 2.3e-66;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 GTTATAGGCTTGAAGTGAACAGTTGGTCAAGATTTGGCAGGAGAGATGGCTTCTT 109

Db 1 GTTATAGGCTTGAAGTGAACAGTTGGTCAAGATTTGGCAGGAGAGATGGCTTCTT 60

Qy 110 CACAAGATGATGAGAGAGAGTGGAGGTGATTTCCAAAGAAACAAGAACCAAGGAA 169

Db 61 CACAAGATGATGAGAGAGAGTGGAGGTGATTTCCAAAGAAACAAGAACCAAGGAA 120

Qy 170 ACACTTCTGAAAAAATTGGGTTTCTCAGAAAT 202

Db 121 ACACTTCTGAAAAAATTGGGTTTCTCAGAAAT 153

RESULT 13
BH618393/c

LOCUS BH618393 443 bp DNA linear GSS 30-JAN-2002
 DEFINITION SALK_030005 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_030005, genomic survey sequence.

ACCESSION BH618393
 VERSION BH618393.1 GI:18428488
 KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 443)
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
 A Sequence-indexed library of insertion mutations in the Arabidopsis Genome Unpublished (2001)

TITLE Arabidopsis thaliana (thale cress)
 JOURNAL Contact: Joseph R. Ecker
 COMMENT Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: eckersalk.edu
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated intron of At5g23570.
 Class: TDNA tagged.

FEATURES
 Location/Qualifiers
 1..443
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone_lib="SALK_030005"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
 Query Match 7.9%; Score 148; DB 28; Length 443;
 Best Local Similarity 100.0%; Pred. No. 1.1e-63;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGGCTGGGATGGGCAACCAAGAGTGTGATATCTGACCAAGTATGAGGCTCTTA 1021
 |||||
 DB 407 AGTGGCTGGGATGGGCAACCAAGAGTGTGATATCTGACCAAGTATGAGGCTCTTA 348
 |||||

QY 1022 GAGCAAGCGCATTCCTATGATGTCACAGAGGCGCATGCGGATGAGTGTTCGATGTTTGAGA 1081
 |||||
 DB 347 GAGCAAGCGCATTCCTATGATGTCACAGAGGCGCATGCGGATGAGTGTTCGATGTTTGAGA 288
 |||||

QY 1082 GCAGTGCCACCTGGCTATTGAGGCCCA 1109
 |||||
 DB 287 GCAGTGCCACCTGGCTATTGAGGCCCA 260
 |||||

RESULT 14
 BH169448 482 bp DNA linear GSS 03-OCT-2001
 LOCUS BH169448
 DEFINITION SALK_001377 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_001377, genomic survey sequence.

ACCESSION BH169448
 VERSION BH169448.1 GI:15904823
 KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 482)
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
 A Sequence-indexed library of insertion mutations in the Arabidopsis Genome Unpublished (2001)

TITLE Arabidopsis thaliana (thale cress)
 JOURNAL Contact: Joseph R. Ecker
 COMMENT Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: eckersalk.edu
 This is single pass sequence recovered from the left border of TDNA.
 Class: TDNA tagged.

FEATURES
 Location/Qualifiers
 1..482
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone_lib="SALK_001377"
 /clone_1ib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
 Query Match 3.0%; Score 57; DB 28; Length 482;
 Best Local Similarity 100.0%; Pred. No. 3.2e-17;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 AGAGAGACAGAACCAACCAAGCAACCTTGGAACCTTGCTCAGAAATT 202
 |||||
 DB 111 AGAGAGACAGAACCAACCAAGCAACCTTGGAACCTTGCTCAGAAATT 167
 |||||

RESULT 15
 AQ964581 232 bp DNA linear GSS 28-JAN-2000
 LOCUS AQ964581
 DEFINITION LERGX20TR LERG Arabidopsis thaliana genomic clone LERGX20, genomic survey sequence.

ACCESSION AQ964581
 VERSION AQ964581.1 GI:6792282
 KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 232)
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utebach, T., Feldblum, T., Liang, F., Creasy, T. and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
 Unpublished (2000)

TITLE Arabidopsis thaliana (thale cress)
 JOURNAL Contact: Xiaoying Lin
 COMMENT The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>
 Seq primer: TR
 Class: shotgun.

FEATURES
 Location/Qualifiers

source

1. .232
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG_ERECTA"
 /db_xref="taxon:3702"
 /clone_lib="LERGX20"
 /note="Organ: Leaf; Vector: pUC19/K; Total genomic DNA was
 sheared to 0.4-0.7 Kbp before ligation."

ORIGIN

Query Match 2.5%; Score 47; DB 28; Length 232;
 Best Local Similarity 100.0%; Pred. No. 3.2e-12;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCTAGGGCTGTCATGCTTAAGGAAAGACGTTCAAGGTGTT 52
 |||||
 DB 134 TTCTAGGGCTGTCATGCTTAAGGAAAGACGTTCAAGGTGTT 180

Search completed: April 9, 2004, 08:16:07
 Job time : 3161.03 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 20:17:10 ; Search time 92.5698 Seconds
(without alignments)
11258.512 Million cell updates/sec

Title: US-10-030-829-2

Sequence: 1 atgagcttcacgagctgctcc.....ttccaatgaagatgatga 1878

Scoring table: OLIGO_NTC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	481	US-09-615-192A-109	Sequence 109, App
C 2	19	1.0	481	US-09-169-789-109	Sequence 109, App
C 3	19	1.0	495	US-08-975-316-27	Sequence 27, Appl
C 4	19	1.0	495	US-09-615-192A-27	Sequence 27, Appl
C 5	19	1.0	495	US-09-169-789-27	Sequence 27, Appl
6	19	1.0	2000	US-09-041-075A-4	Sequence 4, Appl1
7	19	1.0	2000	US-09-041-075A-4	Sequence 4, Appl1
8	19	1.0	8257	US-09-595-684B-30	Sequence 6, Appl1
9	19	1.0	8503	US-09-620-312D-130	Sequence 30, Appl1
10	19	1.0	1664976	US-08-916-421B-1	Sequence 130, App
11	18	1.0	462	US-09-489-039A-5315	Sequence 1, Appl1
12	18	1.0	1254	US-09-322-478-25	Sequence 25, App
13	18	1.0	1829	US-07-668-352C-15	Sequence 15, Appl
14	18	1.0	1829	US-08-474-379C-15	Sequence 15, Appl
15	18	1.0	1829	US-09-146-249A-15	Sequence 15, Appl
16	18	1.0	1829	US-08-206-188B-15	Sequence 15, Appl
17	18	1.0	1829	PCT-US91-02714-15	Sequence 15, Appl
18	18	1.0	1952	US-09-533-029-89	Sequence 89, Appl
C 19	18	1.0	2001	US-09-489-039A-5217	Sequence 89, Appl
C 20	18	1.0	3031	US-08-785-241-2	Sequence 5217, Ap
C 21	18	1.0	7478	US-10-104-966-15	Sequence 15, Appl
22	18	1.0	8906	US-08-826-267-1	Sequence 1, Appl1
23	18	1.0	35524	US-08-923-137-1	Sequence 1, Appl1
24	18	1.0	42571	US-09-810-347-3	Sequence 3, Appl1
25	18	1.0	100848	US-09-556-002-39	Sequence 39, Appl
26	17	0.9	20	US-09-851-896-62	Sequence 62, Appl
C 27	17	0.9	47	US-09-422-978-3322	Sequence 3322, Ap

28	17	0.9	247	4	US-09-411-977-5	Sequence 5, Appl1
C 29	17	0.9	396	4	US-09-107-532A-767	Sequence 767, App
C 30	17	0.9	428	4	US-09-702-705-1214	Sequence 1214, Ap
C 31	17	0.9	428	4	US-09-736-457-1214	Sequence 1214, Ap
C 32	17	0.9	428	4	US-09-614-124B-1214	Sequence 1214, Ap
C 33	17	0.9	428	4	US-09-671-325-1214	Sequence 1214, Ap
34	17	0.9	477	4	US-09-621-976-16220	Sequence 10972, A
35	17	0.9	536	4	US-09-621-976-16220	Sequence 16220, A
36	17	0.9	561	1	US-08-565-386-3	Sequence 3, Appl1
37	17	0.9	576	4	US-09-107-532A-1728	Sequence 1728, Ap
C 38	17	0.9	793	4	US-09-221-017B-940	Sequence 940, App
C 39	17	0.9	1250	3	US-09-018-584A-36	Sequence 36, Appl
C 40	17	0.9	1284	4	US-09-328-352-594	Sequence 594, App
C 41	17	0.9	1290	4	US-09-543-681A-1061	Sequence 1061, Ap
C 42	17	0.9	1323	4	US-09-489-039A-4961	Sequence 4961, Ap
C 43	17	0.9	1401	4	US-09-134-000C-2355	Sequence 2355, Ap
44	17	0.9	1428	2	US-08-360-022-707	Sequence 7, Appl
45	17	0.9	1496	4	US-09-016-434-1127	Sequence 1127, Ap

ALIGNMENTS

```
RESULT 1
US-09-615-192A-109/c
; Sequence 109, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Biokberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003CAU
; CURRENT APPLICATION NUMBER: US/09/615, 192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
; US-09-615-192A-109

Query Match          1.0%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred.No.12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      74 TGGTTACAGGTTTGCAGC 92
Db      470 TGGTTACAGGTTTGCAGC 452

RESULT 2
US-09-169-789-109/c
; Sequence 109, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Biokberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C2
; CURRENT APPLICATION NUMBER: US/09/169,789
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: US 08/713,000
```

EARLIER FILING DATE: 1996-09-11
NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-109

Query Match 1.0%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92
DB 470 TGGTTCAAGGTTTGGCAGG 452

RESULT 3

US-08-975-316-27/c
Sequence 27, Application US/08975316
Patent No. 5952486

GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
APPLICANT: and GRISON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESSES:
ADDRESSES: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:

CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:

NAME: SLENTH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-27

Query Match 1.0%; Score 19; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 4

US-09-615-192A-27/c
Sequence 27, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
TITLE OF INVENTION: Modification of Plant Lignin Content
FILE REFERENCE: 11000.1003C4U
CURRENT APPLICATION NUMBER: US/09/615,192A
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 495
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-27

Query Match 1.0%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 5

US-09-169-789-27/c
Sequence 27, Application US/09169789

Patent No. 6653528
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
TITLE OF INVENTION: Modification of Plant Lignin Content
FILE REFERENCE: 11000.1003C2
CURRENT APPLICATION NUMBER: US/09/169,789
CURRENT FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 08/975,316
EARLIER FILING DATE: 1997-11-21
EARLIER APPLICATION NUMBER: US 08/713,000
EARLIER FILING DATE: 1996-09-11

NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 495
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-27

Query Match 1.0%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 6

US-09-041-075A-4
Sequence 4, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A

APPLICANT: Radding, Jeffrey A
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 2000
TYPE: DNA
ORGANISM: Candida kruisii
FEATURE:
NAME/KEY: CDS
LOCATION: (300)..(1739)
US-09-041-075A-4

Query Match 1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 GATGATTCGATGACGACC 541
DB 1560 GATGATTCGATGACGACC 1578

RESULT 7
US-09-041-075A-6
Sequence 6, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 2000
TYPE: RNA
ORGANISM: Candida kruisii
US-09-041-075A-6

Query Match 1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 73.7%; Pred. No. 12;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 523 GATGATTCGATGACGACC 541
DB 1560 GAUGAUUCGAGACGACC 1578

RESULT 8
US-09-595-684B-30
Sequence 30, Application US/09595684B
Patent No. 6544766
GENERAL INFORMATION:
APPLICANT: Berard, Christophe
APPLICANT: Ohashi, Cara
APPLICANT: Sakowicz, Roman
APPLICANT: Vaisberg, Eugene
APPLICANT: Wood, Kenneth
APPLICANT: Yu, Ming
TITLE OF INVENTION: Human kinesins and methods of producing
TITLE OF INVENTION: and purifying human kinesins
FILE REFERENCE: cytop036

CURRENT APPLICATION NUMBER: US/09/595,684B
CURRENT FILING DATE: 2002-06-24
PRIOR APPLICATION NUMBER: 09/295,612
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 8257
TYPE: DNA
ORGANISM: Human
US-09-595-684B-30

Query Match 1.0%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1802 TTGATCTGGAGAAAGATT 1820
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 9
US-09-620-312D-130
Sequence 130, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghaast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes Version 1.0
SEQ ID NO 130
LENGTH: 8503
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (91)..(8082)
US-09-620-312D-130

Query Match 1.0%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1802 TTGATCTGGAGAAAGATT 1820
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 10

```
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jamaaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jamaaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
```



```
NAME/KEY: misc.feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match 1.0%; Score 19; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Gaps 0;
QY 1702 GAGAAAGAGATGAGAGAT 1720
DB 358045 GAGAAAGAGATGAGAGAT 358063
```

```
RESULT 11
US-09-489-039A-5315
Sequence 5315, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT FILING DATE: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 5315
LENGTH: 462
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5315
```

```
Query Match 1.0%; Score 18; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;
QY 1616 GCCAGCAGCAGCAGACA 1633
DB 216 GCCAGCAGCAGCAGACA 233
```

```
RESULT 12
US-09-322-478-25/c
Sequence 25, Application US/09322478
```

```
Patent No. 6331662
GENERAL INFORMATION:
APPLICANT: Wright, David A.
APPLICANT: Voytas, Daniel F.
TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
FILE REFERENCE: P-1065 ISURF Plant Retroelement
CURRENT FILING DATE: 1999-05-28
EARLIER FILING DATE: 1998-05-29
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 25
LENGTH: 1254
TYPE: DNA
ORGANISM: Pisum sativum
US-09-322-478-25
```

```
Query Match 1.0%; Score 18; DB 4; Length 1254;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;
QY 1335 TCAGCAGCTGACTACTT 1352
DB 890 TCAGCAGCTGACTACTT 873
```

```
RESULT 13
US-07-688-352C-15
Sequence 15, Application US/07688352C
Patent No. 5527896
GENERAL INFORMATION:
APPLICANT: Wiegler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
ADDRESSER: Bicknell.
STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT FILING DATE: 19910419
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET INFORMATION: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1829 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
```

FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-07-688-352C-15

Query Match 1.0%; Score 18; DB 1; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156
Db 985 ATTCCAGAGAACAG 1002

RESULT 14
US-08-474-379C-15
Sequence 15, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/688,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ. ID NO. 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1829 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-08-474-379C-15

Query Match 1.0%; Score 18; DB 2; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156

Db 985 ATTCCAGAGAACAG 1002

RESULT 15
US-09-146-249A-15
Sequence 15, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO. 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1829 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-09-146-249A-15

Query Match 1.0%; Score 18; DB 3; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156
Db 985 ATTCCAGAGAACAG 1002

Search completed: April 9, 2004, 08:20:46
Job time : 96.5698 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 05:52:11 ; Search time 529.543 Seconds
(without alignments)
13304.392 Million cell updates/sec

Title: US-10-030-829-2

Sequence: 1 atgagcttcaggcgtgtcc.....ttccaatgagatgttga 1878

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2475585 seqs, 1875730760 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	1.1	1377	12	US-10-424-599-91177 Sequence 91177, A
2	20	1.1	1589	12	US-10-424-599-91176 Sequence 91176, A
3	19	1.0	172	12	US-10-085-783A-19246 Sequence 19246, A
4	19	1.0	172	12	US-10-242-535A-19246 Sequence 19246, A
5	19	1.0	397	12	US-10-424-599-57837 Sequence 57837, A
6	19	1.0	429	15	US-10-027-632-293476 Sequence 293476, A
7	19	1.0	481	14	US-10-174-693-109 Sequence 109, App
8	19	1.0	495	14	US-10-174-693-27 Sequence 27, App
9	19	1.0	501	9	US-09-833-790-167 Sequence 167, App
10	19	1.0	681	12	US-10-424-599-18549 Sequence 18549, A
11	19	1.0	696	15	US-10-027-632-23555 Sequence 23555, A
12	19	1.0	696	15	US-10-027-632-23556 Sequence 23556, A
13	19	1.0	824	15	US-10-027-632-11811 Sequence 11811, A
14	19	1.0	824	15	US-10-027-632-11812 Sequence 11812, A
15	19	1.0	1137	15	US-10-369-493-24736 Sequence 24736, A

16	19	1.0	1341	9	US-09-938-842A-2465 Sequence 2465, Ap
17	19	1.0	1341	11	US-09-938-842A-2465 Sequence 2465, Ap
18	19	1.0	1576	12	US-10-424-599-4367 Sequence 4367, Ap
19	19	1.0	2000	9	US-09-742-582-4 Sequence 4, Appl
20	19	1.0	2000	9	US-09-742-582-6 Sequence 6, Appl
21	19	1.0	2000	10	US-09-742-580-4 Sequence 4, Appl
22	19	1.0	2000	10	US-09-742-580-6 Sequence 6, Appl
23	19	1.0	2000	10	US-09-742-581-4 Sequence 4, Appl
24	19	1.0	2000	10	US-09-742-581-6 Sequence 6, Appl
25	19	1.0	2240	15	US-10-108-260A-1451 Sequence 1451, Ap
26	19	1.0	3579	14	US-10-176-847-9 Sequence 9, Appl
27	19	1.0	6930	10	US-09-770-107-1 Sequence 1, Appl
28	19	1.0	7992	10	US-09-893-519A-140 Sequence 140, App
29	19	1.0	8491	14	US-10-133-013-260 Sequence 260, App
30	19	1.0	8493	13	US-10-071-766-51 Sequence 51, Appl
31	19	1.0	8503	14	US-10-037-270-130 Sequence 130, App
32	19	1.0	8503	15	US-10-117-722-130 Sequence 130, App
33	19	1.0	47448	15	US-10-085-117-145 Sequence 145, App
34	19	1.0	1691139	14	US-10-067-514-1 Sequence 1, Appl
35	19	1.0	1691139	15	US-10-419-723-1 Sequence 1, Appl
36	19	1.0	2940917	15	US-10-027-632-174763 Sequence 174763, A
37	18	1.0	182	14	US-10-029-386-26108 Sequence 26108, A
38	18	1.0	189	9	US-09-864-761-32008 Sequence 32008, A
39	18	1.0	209	9	US-09-860-352-12756 Sequence 12756, A
40	18	1.0	256	12	US-10-424-599-116017 Sequence 116017, A
41	18	1.0	270	9	US-09-960-352-14374 Sequence 14374, A
42	18	1.0	304	12	US-10-424-599-112786 Sequence 112786, A
43	18	1.0	312	9	US-09-728-445-290 Sequence 290, App
44	18	1.0	315	9	US-09-960-352-12961 Sequence 12961, A
45	18	1.0	395	9	US-09-960-352-13954 Sequence 13954, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-91177
; Sequence 91177, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424, 599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 285684
; SEQ. ID NO 91177
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177

Query Match 1.1%; Score 20; DB 12; Length 1377;
Best local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 673 CAGTGCATTGTCACCTTG 692
DB 359 CAGTGCATTGTCACCTTG 378
RESULT 2
US-10-424-599-91176
; Sequence 91176, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J

```

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91176
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
; US-10-424-599-91176

Query Match          1.1%; Score 20; DB 12; Length 1589;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      673 CAGTGCATTGTCAGCTTG 692
DB      359 CAGTGCATTGTCAGCTTG 378

RESULT 3
US-10-085-783A-19246
; Sequence 19246, Application US/10085783A
; Publication No. US2004003784A1
; GENERAL INFORMATION:
; APPLICANT: Liwew, C.C.
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19246
; LENGTH: 172
; TYPE: DNA
; ORGANISM: Human
; US-10-085-783A-19246

Query Match          1.0%; Score 19; DB 12; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1767 GGAAGACATGAAGAAGAG 1785
DB      86 GGAAGACATGAAGAAGAG 104

RESULT 4
US-10-242-535A-19246
; Sequence 19246, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liwew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
```

```

; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19246
; LENGTH: 172
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-19246

Query Match          1.0%; Score 19; DB 15; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1767 GGAAGACATGAAGAAGAG 1785
DB      86 GGAAGACATGAAGAAGAG 104

RESULT 5
US-10-424-599-57837
; Sequence 57837, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 57837
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_23239C.1
; US-10-424-599-57837

Query Match          1.0%; Score 19; DB 12; Length 397;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1700 AAGAGAAAGATGAGGA 1718
DB      262 AAGAGAAAGATGAGGA 280

RESULT 6
US-10-027-632-293476
; Sequence 293476, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

```
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293476
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-293476
```

```
Query Match          1.0%; Score 19; DB 15; Length 429;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      488 ATGCTTCTGAGAGAGAA 506
DB      76 ATGCTTCTGAGAGAGAA 94
```

```
RESULT 7
US-10-174-693-109/c
; Sequence 109, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Blockberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-109
```

```
Query Match          1.0%; Score 19; DB 14; Length 481;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      74 TGGTCAAGGTTGGCAG 92
DB      470 TGGTCAAGGTTGGCAG 452
```

```
RESULT 8
US-10-174-693-27/c
; Sequence 27, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Blockberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
```

```
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-27
```

```
Query Match          1.0%; Score 19; DB 14; Length 495;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      74 TGGTCAAGGTTGGCAG 92
DB      484 TGGTCAAGGTTGGCAG 466
```

```
RESULT 9
US-09-833-790-167
; Sequence 167, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodch
; APPLICANT: Inditias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-790-167
```

```
Query Match          1.0%; Score 19; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1802 TTGATCTGAGAAAGATT 1820
DB      10 TTGATCTGAGAAAGATT 28
```

```
RESULT 10
US-10-424-599-18549/c
; Sequence 18549, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

SEQ ID NO 18549
LENGTH: 681
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_116754C.1
US-10-424-599-18549

Query Match 1.0%; Score 19; DB 12; Length 681;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1745 TAAAGATCGAGAGAGAA 1763
DB 371 TAAAGATCGAGAGAGAA 353

RESULT 11
US-10-027-632-23555/c
Sequence 23555, Application US/10027632
Publication No. US20030204075A9

GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23555
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23555

Query Match 1.0%; Score 19; DB 15; Length 696;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1122 GGAGTTAGCTGAGATGGGG 1140
DB 82 GGAGTTAGCTGAGATGGGG 64

RESULT 12
US-10-027-632-23556/c
Sequence 23556, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23556
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23556

Query Match 1.0%; Score 19; DB 15; Length 696;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1122 GGAGTTAGCTGAGATGGGG 1140
DB 82 GGAGTTAGCTGAGATGGGG 64

RESULT 13
US-10-027-632-11811
Sequence 11811, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11811
LENGTH: 824
TYPE: DNA
ORGANISM: Human
US-10-027-632-11811

Query Match 1.0%; Score 19; DB 15; Length 824;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1491 CAGGGAAGATGATGCA 1509
DB 579 CAGGGAAGATGATGCA 597

RESULT 14
US-10-027-632-11812
Sequence 11812, Application US/10027632

Job time : 534.543 secs

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11812
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-11812
```

```
Query Match 1.0%; Score 19; DB 15; Length 824;
Best local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1491 CAGGGAAGAGATGATGCA 1509
DB 579 CAGGGAAGAGATGATGCA 597
```

```
RESULT 15
US-10-369-493-24736
; Sequence 24736; Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24736
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; US-10-369-493-24736
```

```
Query Match 1.0%; Score 19; DB 15; Length 1137;
Best local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1702 GAGAAAGAGATGAGAGCT 1720
DB 607 GAGAAAGAGATGAGAGCT 625
```

Search completed: April 9, 2004, 11:58:51

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:35:31 ; Search time 59 Seconds
(without alignments)
2993.088 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625

Sequence: 1 MSSRAGPMSEKXNVQGGYRP.....EPDEALQOLMYKHLNEDD 625

Scoring table: OLIGO

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp1980s:*\n2: geneseqp1390s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	625	100.0	625	4	AAB31798	Aab31798 Amino aci
2	9	1.4	227	4	AA91534	Aag91534 C glutam
3	9	1.4	237	4	AAB79177	Aab79177 Coryneb
4	8	1.3	483	3	AAB18189	Aab18189 Plasmodin
5	8	1.3	916	6	ABU23489	Abu23489 Protein e
6	8	1.3	1553	7	ADC99064	Adc99064 Human KPP
7	8	1.3	1572	5	AAE19162	Aae19162 Human kin
8	7	1.1	15	5	ABG71491	Abg71491 Prostati
9	7	1.1	18	6	ABP82841	Abp82841 G protein
10	7	1.1	32	4	AAE12564	Aae12564 p3serj/CI
11	7	1.1	34	4	AAO04467	Aao04467 Human pol
12	7	1.1	36	3	AAI13241	Aai13241 Caenorhab
13	7	1.1	43	5	AAO18208	Aao18208 Murine ro
14	7	1.1	43	5	AAO18207	Aao18207 Murine ro
15	7	1.1	51	4	AAU43846	Aau43846 Propionib
16	7	1.1	51	6	ABM40365	Abm40365 Propionib
17	7	1.1	57	5	ABP42713	Abp42713 Human ova
18	7	1.1	69	4	AA674546	Aag74546 Human col
19	7	1.1	78	5	ABP08249	Abp08249 Human ORF
20	7	1.1	89	4	AA893334	Aam89334 Human imm
21	7	1.1	90	4	AAU41632	Aau41632 Propionib
22	7	1.1	90	6	ABM38151	Abm38151 Propionib
23	7	1.1	99	5	ABP02735	Abp02735 Human ORF
24	7	1.1	103	4	AAU67643	Aau67643 Propionib
25	7	1.1	103	6	ABM64162	Abm64162 Propionib

26	7	1.1	103	6	ABM64798	Abm64798 Propionib
27	7	1.1	104	4	AAU21876	Aau21876 Human car
28	7	1.1	104	7	ADE45844	Ade45844 Human car
29	7	1.1	108	5	ABG71490	Abg71490 Human pro
30	7	1.1	123	4	AAU66188	Aau66188 Propionib
31	7	1.1	123	6	ABM62707	Abm62707 Propionib
32	7	1.1	124	4	ABG16117	Abg16117 Novel hum
33	7	1.1	138	6	ABP71171	Abp71171 Human pro
34	7	1.1	139	6	ABP71174	Abp71174 Chicken p
35	7	1.1	139	6	ABP71172	Abp71172 Rat pro-N
36	7	1.1	139	6	ABP71173	Abp71173 Mouse pro
37	7	1.1	141	6	ABU23918	Abu23918 Protein e
38	7	1.1	148	4	ABG19350	Abg19350 Novel hum
39	7	1.1	148	8	ABE52649	Ade52649 Murine dn
40	7	1.1	174	4	ABG29097	Abg29097 Novel hum
41	7	1.1	175	5	AAU99286	Aau99286 Rotavir
42	7	1.1	182	4	AAU56433	Aau56433 Propionib
43	7	1.1	182	6	ABM52952	Abm52952 Propionib
44	7	1.1	211	4	AAU76573	Aau76573 Coryneb
45	7	1.1	213	6	ABP56866	Abp56866 Mouse CIT

ALIGNMENTS

RESULT 1	
AA31798	
ID	AA31798 standard; protein; 625 AA.
XX	
AC	AA31798;
XX	
DT	15-MAY-2001 (first entry)
XX	
DE	Amino acid sequence of the Arabidopsis SGS3 polypeptide.
XX	
XX	SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW	viral resistance; resistance; fatty acid content; protein content.
XX	
OS	Arabidopsis thaliana.
XX	
XX	
PN	W0200105951.A2.
XX	
PD	25-JAN-2001.
XX	
PF	13-JUL-2000; 2000MO-FR002052.
XX	
PR	16-JUL-1999; 99FR-00009417.
PR	26-JAN-2000; 2000FR-00001006.
PA	(AVET) AVENTIS CROPS SCIENCE SA.
XX	(INRG) INST NAT RECH AGRONOMIQUE.
XX	
XX	Beclin C, Elmeyan T, Vaucheret H;
XX	
DR	WPI; 2001-159529/16.
XX	
PT	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT	resistance in plants and, when inhibited, for increasing transgene
PT	expression.
XX	
PS	Claim 8; Fig 1; 36pp; French.
XX	
CC	The present sequence represents an Arabidopsis thaliana SGS3 polypeptide.
CC	The SGS3 gene is essential for post-transcriptional inactivation
CC	(degradation of RNA) and for resistance to viruses. Overexpression of
CC	SGS3 results in plants with increased resistance to viruses, while
CC	inactivation of SGS3 in transgenic plants (e.g. by expressing antisense
CC	RNA, by mutation or by homologous recombination) increases the level of
CC	the transgene product. This product may e.g. impart resistance (to
CC	herbicide, insects or pathogens), alter contents of essential fatty acids
CC	or proteins, or is pharmaceutically active, e.g. an immunoglobulin or
CC	interferon

XX Sequence 625 AA; 100.0%; Score 625; DB 4; Length 625;
 SQ Query Match 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSRAGPMSEKKNVGGYRPEVEQIVQGLATGRLASSQDDGGEVYISKKNKPKGNTSG 60
 DB 1 MSSRAGPMSEKKNVGGYRPEVEQIVQGLATGRLASSQDDGGEVYISKKNKPKGNTSG 60
 QY 61 KTVVSGNSNPRAWGQQOQGRGNSVGRGNVSGRNGRGRIQANISGRGRLSRKYDN 120
 DB 61 KTVVSGNSNPRAWGQQOQGRGNSVGRGNVSGRNGRGRIQANISGRGRLSRKYDN 120
 QY 121 NPVAPVSRPPLGGNNQARGSAQHTAVQEPVEDVDVNASSEENDSDALDSDDD 180
 DB 121 NPVAPVSRPPLGGNNQARGSAQHTAVQEPVEDVDVNASSEENDSDALDSDDD 180
 QY 181 LASDYSVDSQSHSGRKNKPKFPGSLDLSLBOINEPOROMHCPACONGPGLIDW 240
 DB 181 LASDYSVDSQSHSGRKNKPKFPGSLDLSLBOINEPOROMHCPACONGPGLIDW 240
 QY 241 YNLHPLAHARTKGRVYKHLRELAETLEKDLQWRGASVIPCSEIYGQWKGLGDEKDYE 300
 DB 241 YNLHPLAHARTKGRVYKHLRELAETLEKDLQWRGASVIPCSEIYGQWKGLGDEKDYE 300
 QY 301 IVPMPVITIMTRLDKDNQKWLQMGNOELLEFDYDKYEAIRAHHSYGPQHRGMSYLMFE 360
 DB 301 IVPMPVITIMTRLDKDNQKWLQMGNOELLEFDYDKYEAIRAHHSYGPQHRGMSYLMFE 360
 QY 361 SSAGCYLAEERLHRELAEMGLDRIAMGKRSMFSGVRYQGLATKODLDIFNOSHQK 420
 DB 361 SSAGCYLAEERLHRELAEMGLDRIAMGKRSMFSGVRYQGLATKODLDIFNOSHQK 420
 QY 421 TRLKELKSYQEMVYKELRQISEDNQNLNYKTKLSKONKAKYLESLTMSKLRRTA 480
 DB 421 TRLKELKSYQEMVYKELRQISEDNQNLNYKTKLSKONKAKYLESLTMSKLRRTA 480
 QY 481 EENRIVROTKMOHONREMDAHDHFPMDSIKQIHERDAKEENFEMLQOQERAKYVQ 540
 DB 481 EENRIVROTKMOHONREMDAHDHFPMDSIKQIHERDAKEENFEMLQOQERAKYVQ 540
 QY 541 OQONINPSSNDCKRKAEEVSSFIFFQEKEMEEFYEEBMLIKOEKCKMCKRHHEI 600
 DB 541 OQONINPSSNDCKRKAEEVSSFIFFQEKEMEEFYEEBMLIKOEKCKMCKRHHEI 600
 QY 601 FDLKEFDEALBQIMYKGLHNEED 625
 DB 601 FDLKEFDEALBQIMYKGLHNEED 625

RESULT 2
 AAG91534 standard; protein; 227 AA.
 ID AAG91534
 XX AAG91534;
 AC AAG91534;
 XX 26-SEP-2001 (first entry)
 DT 26-SEP-2001 (first entry)
 XX C glutamicum protein fragment SEQ ID NO: 5288.
 DE C glutamicum protein fragment SEQ ID NO: 5288.
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX Corynebacterium glutamicum.
 OS Corynebacterium glutamicum.
 XX EPI108790-A2.
 FN EPI108790-A2.
 XX 20-JUN-2001.
 PD 20-JUN-2001.
 XX 18-DEC-2000; 2000EP-00127688.
 PF 18-DEC-2000; 2000EP-00127688.
 XX

PR 16-DEC-1999; 99UP-00377484.
 PR 07-APR-2000; 2000UP-00159162.
 PR 03-AUG-2000; 2000JP-00280988.
 XX
 PA (KYO) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAH66753.
 XX
 PS Claim 17; SEQ ID NO 5288; 246pp + Sequence Listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of Coryneform bacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC Coryneform bacterium. Coryneform bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office

XX Sequence 227 AA;
 SQ Query Match 1.4%; Score 9; DB 4; Length 227;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 DSDALDSD 178
 DB 179 DSDALDSD 187

RESULT 3
 AAB79177 standard; protein; 237 AA.
 ID AAB79177
 XX AAB79177;
 AC AAB79177;
 XX 30-APR-2001 (first entry)
 DT 30-APR-2001 (first entry)
 XX Corynebacterium glutamicum HA protein sequence SEQ ID NO:310.
 DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:310.
 XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
 KW fine chemical production; organic acid; proteiogenic; amino acid;
 KW nonproteinoic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;
 KW Brevibacterium; environmental condition.
 XX Corynebacterium glutamicum.
 OS Corynebacterium glutamicum.
 XX WO200100842-A2.
 FN WO200100842-A2.
 XX 04-JAN-2001.
 PD 04-JAN-2001.
 XX 23-JUN-2000; 2000WO-IB000911.
 PF 23-JUN-2000; 2000WO-IB000911.
 XX 25-JUN-1999; 99US-0141031P.
 PR 08-JUL-1999; 99DE-01031636.
 PR 09-JUL-1999; 99DE-01032125.
 PR 09-JUL-1999; 99DE-01032126.
 PR 09-JUL-1999; 99DE-01032127.
 PR 09-JUL-1999; 99DE-01032127.
 XX

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Bordetella pertussis.

MO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002MO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyckind JW, Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH; WPI; 2003-029926/02.

N-PSDB; ACA27359.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 51413; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for the gene product or that has an activity against a biological pathway; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway; (8) required for proliferation, or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 916 AA;

Query Match 1.3%; Score 8; DB 6; Length 916; Best Local Similarity 100.0%; Pctid. No. 96; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

607 FDEALEQL 614
654 FDEALEQL 661

RESULT 6

ADCG9064 standard; protein; 1553 AA.

ADCG9064;

01-JUN-2004 (first entry)

Human KRP protein - SEQ ID 17.

anti-HIV; anti-allergic; anti-inflammatory; antianaemic; antiparkinsonian; nootropic; anticonvulsant; antiarteriosclerotic; antiscabetic; immunosuppressive; antithyroid; cytosstatic; hepatotropic; dermatological; antidiabetic; nephrotropic; angiotensin; thyromimetic; neuroprotective; osteoplastic; antiarthritic; antiparasitic; antihelminthic; antiparasitic; uroepathic; ophthalmological; antirheumatic; haemostatic; antibacterial; vitricide; protozoacide; fungicide; kinase; phosphatase; KRP; cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; cancer; developmental; mental retardation; neurological; Crohn's; Alzheimer's disease; Parkinson's; autoimmune; inflammatory; Crohn's; diabetes mellitus; viral; bacterial; fungal; parasitic; protozoan; helminthic infection; transgenic; gene therapy; human; enzyme.

Homo sapiens.

MO2003033680-A2.

24-APR-2003.

17-OCT-2002; 2002MO-US033723.

19-OCT-2001; 2001US-0345474P.

02-NOV-2001; 2001US-0343910P.

13-NOV-2001; 2001US-033098P.

16-NOV-2001; 2001US-0332424P.

30-NOV-2001; 2001US-0334288P.

(INCY-) INCYTE GENOMICS INC.

Bandman O, Baughn MR, Becha SD, Borowsky ML, Duggan BM; Emerling BW, Forsythe J, Gandhi AR, Gorvad AE, Griffin JA; Gururajan R, Hafalia AJA, Khan PA, Lal PG, Lee EA, Lee ST; Gururajan R, Lu DM, Lu Y, Marquis JP, Nguyen DB, Arvizu CS; Ramkumar J, Recipon SA, Richardson TW, Swarnakar A, Tang YT; Thornton MB, Tran UK, Chawla NK, Warren BA, Yang J, Yao MG, Yue H; Zebajradan Y;

WPI; 2003-403214/38.

N-PSDB; ADC99116.

New human kinases and phosphatases and polynucleotides, useful for diagnosing, treating or preventing autoimmune or inflammatory disorders (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis, cancer or hepatitis.

Claim 1; SEQ ID NO 17; 424pp; English.

The invention relates to a novel isolated polypeptide which is a human kinase and phosphatase (KRP). The KRP polypeptides, polynucleotides, agonists and antagonists are useful for diagnosing, treating or preventing cell proliferative disorders such as atherosclerosis, cirrhosis, hepatitis and cancer, developmental disorders e.g. mental retardation, neurological disorders including Alzheimer's disease and Parkinson's disease, autoimmune and inflammatory disorders such as Crohn's disease and diabetes mellitus and finally, viral, bacterial, fungal, parasitic, protozoan or helminthic infections. Furthermore, the polynucleotides encoding KRP may be useful for creating transgenic animals to model human disease, as well as during gene therapy procedures. The current sequence is that of the human KRP protein of the invention.

Sequence 1553 AA;

Query Match 1.3%; Score 8; DB 7; Length 1553;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 RLHRELA 378
 |||||
 Db 493 RLHRELA 500

RESULT 7
 AAE19162
 ID AAE19162 standard; protein; 1572 AA.
 AC AAE19162;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Human kinase polypeptide (PKIN-20).
 XX
 KW Human; kinase polypeptide; PKIN-20; gene therapy; Addison's disease;
 KW leukemia; immune disorder; lymphoma; melanoma; developmental disorder;
 KW acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hyperemion;
 KW asthma; Crohn's disease; rheumatoid arthritis; burstlitis; atherosclerosis;
 KW cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cystostatic; cancer;
 KW cholestasis; anti-HIV; cardiovascular disorder; Niemann-Pick's disease;
 KW lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;
 KW drug screening; transgenic animal; antiinflammatory; hepatotropic;
 KW hypotensive; cardiac; enzyme.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..37
 FT Protein /label= Signal_peptide
 FT Domain /note= "Mature human PKIN-20"
 FT Binding-site /note= "Eukaryotic proetin kinase domain"
 FT Binding-site /note= "Phorbol esters/diacylglycerol binding domain"
 FT Binding-site /note= "900..963
 FT Domain /note= "956..1074
 FT Domain /note= "8CNH domain"
 FT Domain /note= "1100..1380
 FT Domain /note= "CNH domain"
 XX
 PN WO200208399-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 20-JUL-2001; 2001WO-US023092.
 XX
 PR 21-JUL-2000; 2000US-0220038P.
 PR 28-JUL-2000; 2000US-0222112P.
 PR 04-AUG-2000; 2000US-0222831P.
 PR 11-AUG-2000; 2000US-0224729P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PA (THOR/) THORNTON M.
 XX
 PI Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AD, Walia NK;
 PI Patterson C, Ramkumar J, Gandhi AR, Policky JL, Baughn MR,
 PI Tribouley CM, Bandman O, Nguyen DB, Lu Y, Buford N, Lai P, Ding L,
 PI Yao MG, Elliott VS, Recipon SA, Kearney L, Lu DM, Greenwald SR;
 PI Tang YT, Xu Y, Walsh RT, Gietzen KU, Yang J, Hillman JL;
 XX
 DR WPI; 2002-206083/26.
 DR N-PSDB; AAD30567.
 XX
 PT New human kinase polypeptide, useful in diagnosis, prevention and
 PT treatment of cancer, immune disorder, growth and developmental disorder,
 PT cardiovascular disorder and lipid disorder.

XX
 PS Claim 1; Page 170-174; 196pp; English.
 XX
 CC The present invention relates to an isolated human kinase polypeptide
 CC (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is
 CC useful for diagnosing, treating and preventing cancer (e.g., leukaemia,
 CC lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency
 CC syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's
 CC disease, rheumatoid arthritis), a growth and developmental disorder (e.g.
 CC burstlitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a
 CC cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial
 CC infarction), and a lipid disorder (e.g., fatty liver, cholestasis,
 CC Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of
 CC drug screening techniques and to analyse the proteome of a tissue or cell
 CC type. PKIN is useful for creating knockin humanised animals or transgenic
 CC animals to model human diseases, in somatic or germ-line gene therapy, to
 CC generate a transcript image of a tissue or cell type, for detecting
 CC differences in the chromosomal location due to translocation, inversion,
 CC etc., among normal, carrier or affected individuals, and as hybridisation
 CC probes for mapping naturally occurring genomic sequences. PKIN is useful
 CC in southern or northern analysis, dot blot or other membrane-based
 CC technologies, in PCR technologies, in dipstick, pin, multiformat enzyme
 CC linked immunosorbent (ELISA)-like assays and in microarrays utilising
 CC fluids or tissues from patients to detect altered PKIN expression. The
 CC present sequence is human PKIN-20
 XX
 SO Sequence 1572 AA;

QY 371 RLHRELA 378
 |||||
 Db 493 RLHRELA 500

RESULT 8
 ABG71491
 ID ABG71491 standard; peptide; 15 AA.
 XX
 AC ABG71491;
 XX
 DT 28-FEB-2003 (first entry)
 XX
 DE Prostatic specific membrane antibody protein 11.88 N-terminal peptide.
 XX
 KW Human; prostatic specific membrane antibody protein 11.88; antibody;
 KW prostatic cancer; prostatic benign tumour; folic acid; metabolism;
 KW haemopathy; human immunodeficiency virus; HIV; immunological disease;
 KW inflammation.
 XX
 OS Homo sapiens.
 XX
 PN CN1352107-A.
 XX
 PD 05-JUN-2002.
 XX
 PR 06-NOV-2000; 2000CN-00127237.
 PR 06-NOV-2000; 2000CN-00127237.
 XX
 PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
 PA Mao Y, Xie Y;
 XX
 DR WPI; 2002-692403/75.
 XX
 PT New human prostatic specific membrane antibody protein 11.88 polypeptide
 PT for treating e.g. prostatic cancer, folic acid metabolic fault, human
 PT immunodeficiency virus infection, immunological diseases, and
 PT inflammations.
 XX

PS Example 5; Page 19 (disclosure); 33pp; Chinese.
 CC The invention relates to the human prostatic specific membrane antibody
 CC protein 11.88, a polynucleotide encoding the polypeptide and a DNA
 CC recombinant process used to produce the polypeptide. The polypeptide
 CC and the polynucleotide are used for treating various diseases, such as
 CC prostatic cancer, prostatic benign tumour, other tumours, folic acid
 CC metabolism fault, haemopathy, human immunodeficiency virus (HIV)
 CC infection, immunological diseases and inflammations. This sequence
 CC represents a human prostatic specific membrane antibody protein 11.88 N-
 CC terminal peptide, used in ELISA
 XX Sequence 15 AA;
 SQ
 QY Query Match 1.1%; Score 7; DB 5; Length 15;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 48 SKKNKK 54
 7 SKKNKK 13
 AC ABB2841 standard; peptide; 18 AA.
 XX ABB2841;
 DT 04-MAR-2003 (first entry)
 DE G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:1514.
 XX G protein-coupled receptor (GPCR) antigenic peptide; gene therapy;
 XX G protein-coupled receptor modulator; antibody; immune-related disease;
 XX growth-related disease; cell regeneration-related disease; AIDS; cancer;
 XX immunological-related cell proliferative disease; autoimmune disease;
 XX Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
 XX osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
 XX graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
 XX psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
 XX mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
 XX hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
 XX ulcer.
 XX Homo sapiens.
 XX OS
 XX PN WO200261087-A2.
 XX PD 08-AUG-2002.
 XX PF 19-DEC-2001; 2001WO-US050107.
 XX PR 19-DEC-2000; 2000US-0257144P.
 XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.
 XX PI Burmer GC, Roush CL, Brown JP;
 XX WPI; 2003-046718/04.
 XX DR
 XX PT New isolated antigenic peptides e.g., for G protein-coupled receptors
 XX (GPCR), useful for diagnosing and designing drugs for treating conditions
 XX in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
 XX autoimmune diseases.
 XX PS Claim 1; Fig 2; 523pp; English.
 CC The present invention describes antigenic peptides (1) comprising: (a)
 CC any one of 1601 sequences (see ABB2019 to ABB3619) of 12-24 amino
 CC acids. Also described: (1) an assay for the detection of a particular G
 CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
 CC and (2) an isolated antibody having high specificity and high affinity or

CC avidity for a particular GPCR. (1) can be used as GPCR modulators and in
 CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
 CC antibody against a particular GPCR, and in the production of specific
 CC antibodies. The peptides and antibodies are also useful for detecting the
 CC presence or absence of corresponding GPCRs. The antigenic peptides for
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for
 CC treating immune-related diseases, growth-related diseases, cell
 CC regeneration-related disease, immunological-related cell proliferative
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
 CC any other disorder in which GPCRs are involved. The antibodies may be
 CC used in immunoassays and immunodiagnoses. ABB2523 to ABB24869 encode
 CC GPCR proteins given in ABB1675 to ABB2018, which are used in the
 CC exemplification of the present invention
 XX Sequence 18 AA;
 SQ
 QY Query Match 1.1%; Score 7; DB 6; Length 18;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 176 DSDDDLA 182
 3 DSDDDLA 9
 AC ABB12564 standard; peptide; 32 AA.
 XX ABB12564;
 DT 15-JAN-2002 (first entry)
 DE p35srj/CITBD2 p300-CH1 interacting domain.
 XX Human; transcription transactivator protein; CREB-binding protein;
 XX CITED family; CBP/p300 interacting transactivators with BD-rich tails;
 XX HCITEDX protein; hypoxia signalling; inflammation; tumour angiogenesis;
 XX cholesterol biosynthesis; ischemic heart disease; Netherton's disease;
 XX angiogenesis; atopic disease; cytostatic; vasotropic; gene therapy;
 XX p35srj/CITBD2 p300-CH1 interacting domain.
 XX Homo sapiens.
 XX OS
 XX PN WO200168846-A2.
 XX PD 20-SEP-2001.
 XX PF 16-MAR-2001; 2001WO-GB001201.
 XX PR 17-MAR-2000; 2000GB-00006572.
 XX PA (ISIS-) ISIS INNOVATION LTD.
 XX PI Bhattacharya S, Braganca J, Swingle T;
 XX WPI; 2001-596907/67.
 XX DR
 XX PT New transcription transactivator protein of CITED family, termed HCITEDX,
 XX useful for controlling hypoxia signaling, inflammation, activating
 XX cholesterol uptake genes and identifying compounds interfering with
 XX HCITEDX function.
 XX PS Example 1; Page 50; 95pp; English.
 CC The patent discloses transcription transactivator protein, which is a

CC member of CITED (CREB-binding protein (CBP)/p300 interacting
 CC transactivators with ED-rich tails) family, designated as HCTEDX protein
 CC and its corresponding polynucleotides. HCTEDX proteins are useful for
 CC identifying compounds which modulate the cytoplasmic sequestration of the
 CC HCTEDX protein. They are useful in the control of hypoxia signalling and
 CC in the activation of genes involved in cholesterol uptake, cholesterol
 CC biosynthesis and in the control of inflammation. They are therapeutically
 CC useful in number of different indications, particularly reduction of
 CC cholesterol biosynthesis and inflammation, prevention of tumour
 CC angiogenesis and treatment of any other condition. HCTEDX proteins are
 CC useful in screening methods for identification of compounds which
 CC interfere with the function of HCTEDX and useful for modulating hypoxia
 CC signalling, tumour angiogenesis or cholesterol synthesis. HCTEDX
 CC modulators are useful for preventing tumour angiogenesis and treating of
 CC ischaemic heart disease. Compounds which promote the translocation of
 CC cytoplasmic HCTEDX in breast cancer cells are useful as leads in the
 CC development of anti-breast cancer drugs and compounds which prevent the
 CC translocation of HCTEDX into the nucleus enhance angiogenesis. HCTEDX
 CC modulators are useful for treating Netherton's disease or atopic disease.
 CC HCTEDX sequences are also used in gene therapy. The present peptide
 CC sequence is the p300-CH1 interacting domain of p35serj/CITED2. This
 CC sequence is used to clone full length HCTEDX

SQ Sequence 32 AA;

Query Match 1.1%; Score 7; DB 4; Length 32;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 EMGLDRI 384
 |||||
 Db 11 EMGLDRI 17

RESULT 11
 AAO04467
 ID AAO04467 standard; protein; 34 AA.

XX AAO04467;

AC 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 18359.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US004927.

PR 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI, 2001-514838/56.

DR N-PSDB; AA184398.

PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 and treating e.g. leukemia, inflammation and immune disorders.

PS Claim 20, SEQ ID NO 18359; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, hematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at http://wipo.int/pub/published_pct_sequences

SQ Sequence 34 AA;

Query Match 1.1%; Score 7; DB 4; Length 34;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 SKKNKK 54
 |||||
 Db 23 SKKNKK 29

RESULT 12
 AAB13241
 ID AAB13241 standard; peptide; 36 AA.

XX AAB13241;

DT 11-JAN-2001 (first entry)

DE Caenorhabditis elegans conserved sequence #5.

KW Caenorhabditis elegans; metabolic enzyme; AKT kinase; daf-18;

KW insulin signalling pathway; daf-2; age-1; insulin receptor; pi 3-kinase;

KW PKB kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;

KW diabetes.

XX Caenorhabditis elegans.

PN WO200033068-A1.

PD 08-JUN-2000.

PF 02-DEC-1999; 99WO-US028529.

PR 03-DEC-1998; 98US-00205658.

PA (GEHO) GEN HOSPITAL CORP.

PI Ruvkun G, Ogg S;

DR WPI; 2000-423022/36.

PT Diagnosing and treating obesity and impaired glucose tolerance using
 modulators of daf-18 expression and/or activity.

PS Disclosure; Page 167; 402pp; English.

CC The present sequence is conserved protein region of a Caenorhabditis
 CC elegans homologue of a key metabolic enzyme. A number of C. elegans genes
 CC have been identified as homologues of genes in the mammalian insulin
 CC signalling pathway. The C. elegans age-1 gene encodes a homologue of the
 CC mammalian pi 3-kinase whilst daf-2 encodes a homologue of the mammalian
 CC insulin receptor. The C. elegans AKT kinase and PKB kinase act downstream
 CC of daf-2 and age-1, just as their mammalian homologues act downstream of
 CC insulin signalling. The C. elegans PTEN lipid phosphatase homologue, DAF-
 CC 18, has been found to act upstream of AKT in the pathway. This discovery
 CC has enabled mammalian PTEN action to be mapped to the insulin signalling
 CC pathway. Conserved DAF motifs can be used to design probes to identify
 CC mammalian DAF homologues and thus to identify individuals with a
 CC predisposition toward the development of glucose intolerance conditions,

CC such as obesity and diabetes
XX
SQ Sequence 36 AA;

Query Match 1.1%; Score 7; DB 3; Length 36;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 LROI SED 444
DB 2 LROI SED 8

RESULT 13

AAO18208 standard; peptide; 43 AA.

XX AAO18208;

DT 13-SEP-2002 (first entry)

XX Murine rotavirus non-structural protein 4 NSP4 fragment #2.

DE Rotavirus; non-structural protein 4; NSP4; protein coordinate data;
XX Rotavirus; non-structural protein 4; NSP4; protein coordinate data;
KM crystal structure; antiviral; vaccine; diarrhoea; gastroenteritis.

XX Murine rotavirus.

XX WO200244330-A2.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044555.

XX 01-DEC-2000; 2000US-0250810P.

XX (AMHP) AMERICAN HOME PROD CORP.

XX Bowman GD, Levy O, Nodelman I, Schutt CE, Zamb TJ;

XX WPI; 2002-537448/57.

XX Novel crystal of oligomerization domain of rotavirus nonstructural
PT protein 4, useful for determining 3 dimensional crystal structure of the
PT domain that is used for identifying agents which interact with the
PT domain.

XX Disclosure; Fig 1B; 81pp; English.

XX The present invention provides the crystal structure of the

CC oligomerization domain of the rotavirus non-structural protein 4 (NSP4).

CC Rotavirus is a major cause of gastroenteritis, and the structure can be

CC used to construct peptide useful in vaccines to protect against

CC infection. The present sequence is a fragment of a rotavirus NSP4 protein

XX Sequence 43 AA;

QY 434 VVKELRQ 440

DB 8 VVKELRQ 14

RESULT 14

AAO18207 standard; peptide; 43 AA.

XX AAO18207;

DT 13-SEP-2002 (first entry)

XX Murine rotavirus non-structural protein 4 NSP4 fragment #1.

DE Rotavirus; non-structural protein 4; NSP4; protein coordinate data;
XX Rotavirus; non-structural protein 4; NSP4; protein coordinate data;
KM crystal structure; antiviral; vaccine; diarrhoea; gastroenteritis.

XX Murine rotavirus.

XX WO200244330-A2.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044555.

XX 01-DEC-2000; 2000US-0250810P.

XX (AMHP) AMERICAN HOME PROD CORP.

XX Bowman GD, Levy O, Nodelman I, Schutt CE, Zamb TJ;

XX WPI; 2002-537448/57.

XX Novel crystal of oligomerization domain of rotavirus nonstructural
PT protein 4, useful for determining 3 dimensional crystal structure of the
PT domain that is used for identifying agents which interact with the
PT domain.

XX Disclosure; Fig 1B; 81pp; English.

XX The present invention provides the crystal structure of the

CC oligomerization domain of the rotavirus non-structural protein 4 (NSP4).

CC Rotavirus is a major cause of gastroenteritis, and the structure can be

CC used to construct peptide useful in vaccines to protect against

CC infection. The present sequence is a fragment of a rotavirus NSP4 protein

XX Sequence 43 AA;

QY 434 VVKELRQ 440

DB 8 VVKELRQ 14

RESULT 15

AAU43846 standard; protein; 51 AA.

XX AAU43846;

DT 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #4742.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

PA (CORI-) CORIXA CORP.

XX Skeily YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX

DR WPI; 2001-616774/71.
DR N-PSDB; AAS59521.

XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.

XX
PS Example 1; SEQ ID NO 5041; 1069pp; English.

XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 51 AA;

Query Match 1.1%; Score 7; DB 4; Length 51;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 SGRGN 99

DB 34 SGRGN 40

Search completed: April 6, 2004, 19:42:00
Job time : 62 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:39:52 ; Search time 20 Seconds
(without alignments)
3005.985 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625
Sequence: 1 MSSRAGPMSEKKNVQGGYRP.....EFDEALQQLMYKHGLHND 625

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR 78: **
2: pir1: **
3: pir2: **
4: pir3: **

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.4	603	2 T27901	hypothetical prote
2	8	1.3	353	2 T33782	hypothetical prote
3	8	1.3	483	2 F71619	hypothetical prote
4	8	1.3	586	2 T19406	hypothetical prote
5	7	1.1	69	2 G84303	snRNP homolog (imp
6	7	1.1	101	2 D84164	hypothetical prote
7	7	1.1	135	2 B70456	thioredoxin - Agui
8	7	1.1	141	2 C97338	hydroxymyristoyl-(
9	7	1.1	145	2 A12869	NTP pyrophosphohyd
10	7	1.1	147	1 S24310	flavodoxin - Desul
11	7	1.1	147	2 C97646	hypothetical prote
12	7	1.1	148	1 OTBY6	cytochrome-c oxida
13	7	1.1	149	2 D69831	conserved hypotet
14	7	1.1	152	2 G96010	hypothetical expor
15	7	1.1	182	2 T15118	hypothetical prote
16	7	1.1	207	2 T49552	hypothetical prote
17	7	1.1	212	2 C84007	negative regulator
18	7	1.1	215	2 S77663	multicopy phage re
19	7	1.1	253	2 H72384	conserved hypotet
20	7	1.1	257	2 C40304	neurotrophin-3 pre
21	7	1.1	257	2 F50400	neurotrophin-3 pre
22	7	1.1	257	2 F75476	probable utrophin
23	7	1.1	258	2 S09155	neurotrophin-3 pre
24	7	1.1	268	2 H72483	hypothetical prote
25	7	1.1	271	2 I39491	hypothetical prote
26	7	1.1	272	2 A71710	hypothetical prote
27	7	1.1	275	2 H87546	enoyl-CoA hydratase
28	7	1.1	277	2 G71822	crRNA delta(2)-18op
29	7	1.1	277	2 E97703	hypothetical prote

30	7	1.1	278	2 T30450	late expression fa
31	7	1.1	279	2 S67670	hypothetical prote
32	7	1.1	282	2 A35781	hippocampus-derive
33	7	1.1	283	2 B85700	hypothetical prote
34	7	1.1	283	2 A90843	ychb protein [semi
35	7	1.1	283	2 B47706	ychb protein - Esc
36	7	1.1	285	2 T41656	vacuolar ATP synth
37	7	1.1	286	2 B41810	transcription fact
38	7	1.1	286	2 B82564	acetylxylian estera
39	7	1.1	301	2 JS0681	hypothetical prote
40	7	1.1	301	2 B85087	probable phosphogl
41	7	1.1	304	2 D82304	glutamy1-tRNA synt
42	7	1.1	309	2 F97168	nucleoside-diphosp
43	7	1.1	320	2 S69547	transcription init
44	7	1.1	321	2 AH0412	probable glutamyl-
45	7	1.1	324	2 T00939	hypothetical prote

ALIGNMENTS

RESULT 1
T27901
hypothetical protein ZK546.13 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27901
R:Hallsworth, K.
Submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid ZK546.
A:Reference number: Z20437
A:Accession: T27901
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-603 <HAL>
A:Cross-references: EMBL:U29380; PIDN:AAA68747.1; CESP:ZK546.13
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:ZK546.13
A:Introns: 26/3; 111/3; 196/3; 275/3; 331/1; 362/3; 438/3; 500/3

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 LEIMSEKLR 477
DB 433 LEIMSEKLR 441

RESULT 2
T33782
hypothetical protein C39F7.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C:Accession: T33782
R:Maggi, L.; Scheet, P.; Dubbelde, C.
Submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid C39F7.
A:Reference number: Z21407
A:Accession: T33782
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-353 <MAG>
A:Cross-references: EMBL:AF101310; PIDN:AAC69214.1; GSPDB:GN00023; CESP:C39F7.5
A:Experimental source: strain Bristol N2; clone C39F7
C:Genetics:
A:Gene: CESP:C39F7.5
A:Map position: 5
A:Introns: 14/2; 45/3; 224/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C39F7.5

Query Match 1.3%; Score 8; DB 2; Length 353;

Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 EYEOLVQV 28
DB 342 EYEOLVQV 349

RESULT 3

F71619
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: F71619
C:Author: M.J.; Tetteelin, H.; Canucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
R.Gardner, M.J.; Salzman, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 287, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: F71619
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-483 <GAR>
A:Cross-references: GB:AE001382; GB:AE001362; NID:93845130; PIDN:AACT1836.1; PID:9384513
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0235w

Query Match

Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ISKKNK 54
DB 33 ISKKNK 40

RESULT 4

T19406
hypothetical protein C18E9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T19406
R:Sim, M.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19120
A:Accession: T19406
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-586 <WIL>
A:Cross-references: EMBL:Z70034; PIDN:CA93857.1; GSPDB:GN00020; CESP:C18E9.8
A:Experimental source: clone C18E9
C:Genetics:
A:Gene: CESP:C18E9.8
A:Map position: 2
A:introns: 158/3; 269/3; 354/3; 493/2; 538/2
A:Superfamily: Caenorhabditis elegans hypothetical protein C18E9.8

Query Match

Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 RELAEVLE 269
DB 495 RELAEVLE 502

RESULT 5

G84303
snRNP homolog (imported) - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84303

R.N.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
Jung, K.H.; Alam, M.; Freitas, T. 12176-12181, 2000
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Lie
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: G84303
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-69 <STO>
A:Cross-references: GB:AE004437; NID:910580991; PIDN:AAI19795.1; GSPDB:GN00138
C:Genetics:
A:Gene: snp

Query Match 1.1%; Score 7; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 VLERSLE 470
DB 17 VLERSLE 23

RESULT 6

D84164
hypothetical protein Vng0041c (imported) - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: D84164
R:Ng, M.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Lie
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: D84164
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-101 <STO>
A:Cross-references: GB:AE004437; NID:910579692; PIDN:AAI18680.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0041C
C:Superfamily: Halobacterium plasmid pNRC100 hypothetical protein H0011

Query Match 1.1%; Score 7; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 AHARTG 254
DB 55 AHARTG 61

RESULT 7

B70456
thioredoxin - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C:Accession: B70456
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: B70456
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-135 <AGP>
A:Cross-references: GB:AE000757; NID:92984092; PIDN:AACT635.1; PID:92984097; GB:AE00065
A:Experimental source: strain VFS
C:Genetics:

A:Gene: trxA2

Query Match 1.1%; Score 7; DB 2; Length 135;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 KFFGSID 212
 |||||
 DB 109 KFFGSID 115

RESULT 8

C97338

hydroxymyristoyl-[acyl carrier protein] dehydratase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: C97338

R:Noilling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C

A:Reference number: A96900; MWID:21359325; PMID:21359325

A:Accession: C97338

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-141 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81494.1; PID:gl5026667; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3571

C:Superfamily: (3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase

Query Match 1.1%; Score 7; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 SLSTEQI 219
 |||||
 DB 2 SLSTEQI 8

RESULT 9

A12869

NTP pyrophosphohydrolase, Mutr family [imported] - Agrobacterium tumefaciens (strain C58

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: A12869

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

Science 294, 2317-2323, 2001

A:Reference number: A97359; MWID:21608551; PMID:11743194

A:Accession: A12869

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <KUR>

A:Cross-references: GB:AE008668; PIDN:AA143375.1; PID:gl7740871; GSPDB:GN00186

A:Experimental source: strain C58 (Dugont)

C:Genetics:

A:Gene: Atu2387

A:Map position: circular chromosome

QY 260 LRRELA 266
 |||||
 DB 51 LRRELA 57

Query Match 1.1%; Score 7; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10

S24310

flavodoxin - Desulfovibrio gigas (ATCC 29494)

C:Species: Desulfovibrio gigas

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 05-May-2000

C:Accession: S24310

R:Helm, L.R.; Swenson, R.P.

Biochim. Biophys. Acta 1131, 325-328, 1992

A>Title: The primary structures of the flavodoxins from two strains of Desulfovibrio giga

A:Reference number: S24310; MWID:92329549; PMID:1627649

A:Accession: S24310

A:Molecule type: DNA

A:Residues: 1-147 <HEU>

A:Cross-references: EMBL:664765; NID:940798; PIDN:CAA46012.1; PID:940799

A:Experimental source: strain ATCC 29494

C:Superfamily: Flavodoxin; Flavodoxin homology

C:Keywords: electron transfer; Flavoprotein; FMN

F/6-143/Domain: flavodoxin homology <FLX>

QY 462 AKVLEES 468
 |||||
 DB 23 AKVLEES 29

Query Match 1.1%; Score 7; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11

C97646

hydroxymyristoyl-[acyl carrier protein] dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Cerec

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: C97646

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MWID:21608551; PMID:11743194

A:Accession: C97646

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-147 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK8124.1; PID:gl5157558; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C_4330

A:Map position: circular chromosome

QY 260 LRRELA 266
 |||||
 DB 53 LRRELA 59

Query Match 1.1%; Score 7; DB 2; Length 147;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12

OYEVE

cytochrome-c oxidase (EC 1.9.3.1) chain VI precursor - Yeast (Saccharomyces cerevisiae)

N:Alternate names: protein H8179.2; protein YHR051W

C:Species: Saccharomyces cerevisiae

C>Date: 13-Jun-1983 #sequence_revision 31-Dec-1992 #text_change 23-Mar-2001

C:Accession: A22853; S46730; A00494; S48874

R:Wright, R.M.; Ko, C.; Cumsky, M.G.; Poyton, R.O.

J. Biol. Chem. 259, 15401-15407, 1984

A>Title: Isolation and sequence of the structural gene for cytochrome c oxidase subunit I

A:Reference number: A92466; MWID:85080033; PMID:6210289

A:Accession: A22853

A:Molecule type: DNA

A:Residues: 1-148 <WRI>

A:Cross-references: EMBL:M10138; NID:gl71294; PIDN:AAA66900.1; PID:gl71295

Query Match 1.1%; Score 7; DB 2; Length 182;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 GRGNNG 100
 |||||
 Db 173 GRGNNG 179

Search completed: April 6, 2004, 19:44:01
 Job time : 22 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:36:06 ; Search time 18 Seconds

(without alignments)
1807.991 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625
Sequence: 1 MSSRAGPMSEKKNVQGGYRP.....EFDEALEQLMYKHLNEDD 625

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.4	227	1 NUSB CORGL	Q8ngq3 corynebacte
2	8	1.3	113	1 RBFA OCEIH	Q8egq3 oceanobacti
3	7	1.1	141	1 FABZ CLOAB	Q97da9 clostridium
4	7	1.1	147	1 FLAW DESGI	001096 desulfotrib
5	7	1.1	148	1 COX6 YEAST	P00427 saccharomyc
6	7	1.1	153	1 G6PD SARBU	Q25537 sarcophaga
7	7	1.1	154	1 NT3 CEREL	Q25150 cervus elap
8	7	1.1	242	1 Y370 THEMA	Q9wyk4 thermotoga
9	7	1.1	257	1 NT3 CHICK	P25423 gallus gall
10	7	1.1	257	1 NT3 FELCA	Q9tser2 felis silve
11	7	1.1	257	1 NT3 HUMAN	P20783 homo sapien
12	7	1.1	258	1 NT3 MOUSE	P20181 mus musculu
13	7	1.1	258	1 NT3 RAT	P18280 rattus norv
14	7	1.1	269	1 CIT2 MOUSE	Q35740 mus musculu
15	7	1.1	270	1 CIT2 HUMAN	Q99677 homo sapien
16	7	1.1	271	1 YNR3 AZOBR	P45674 azospirillum
17	7	1.1	283	1 ISPE_ECOL6	Q8fio4 escherichia
18	7	1.1	283	1 ISPE_ECOL1	P24209 escherichia
19	7	1.1	285	1 VATD_SCHPO	O59823 schistosach
20	7	1.1	286	1 TOAI YEAST	P32773 saccharomyc
21	7	1.1	301	1 YGL4 BACST	P32814 bacillus st
22	7	1.1	308	1 CG17 YEAST	P25693 saccharomyc
23	7	1.1	311	1 MIAA_HELPY	O25961 helicobacte
24	7	1.1	312	1 MIAA_HELPY	Q2zj17 helicobacte
25	7	1.1	347	1 ITP2 HUMAN	Q9ukp3 homo sapien
26	7	1.1	360	1 SYFA_AGR15	Q9u1k5 agrobacteri
27	7	1.1	360	1 SYFA_RHIME	Q928r0 rhizobium m
28	7	1.1	372	1 ALR_SYNY3	P54918 synecocyst
29	7	1.1	372	1 ISPG_VIRBA	Q87e16 vibrio para
30	7	1.1	372	1 ISPG_VIRBA	Q8de8 vibrio vuln
31	7	1.1	390	1 TAL_PROMM	Q7v658 prochloroco
32	7	1.1	399	1 T7S1 HUMAN	O60478 homo sapien
33	7	1.1	420	1 TRPB_HELHP	Q7vga7 helicobacte

34	7	1.1	453	1 SR54_ECOLI	P07019 escherichia
35	7	1.1	461	1 ARLY_SYNY3	P73257 synecocyst
36	7	1.1	476	1 YG32 YEAST	P53270 saccharomyc
37	7	1.1	483	1 APM3 YEAST	P38153 saccharomyc
38	7	1.1	510	1 EMR2_HAEIN	P44927 haemophilus
39	7	1.1	523	1 UDBG_RABIT	O19103 oryctolagus
40	7	1.1	527	1 UDAL RAT	P36510 rattus norv
41	7	1.1	528	1 UDB4 HUMAN	P36513 homo sapien
42	7	1.1	528	1 UDBA HUMAN	P36537 homo sapien
43	7	1.1	528	1 UDB3_MACFA	O9x55 macaca fasc
44	7	1.1	529	1 UDB3_MACFA	O02663 macaca fasc
45	7	1.1	529	1 UDBB_HUMAN	O75310 homo sapien

ALIGNMENTS

```

RESULT 1
NUSB CORGL          STANDARD;          PRT;          227 AA.
ID  NUSB CORGL
AC  Q8N0J3;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DE  N utilization substance protein B homolog (NusB protein).
GN  NUSB OR CGL1618.
OS  Corynebacterium glutamicum (Brevibacterium flavum).
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX  NCBI_Taxid=1718;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA  Nakagawa S.;
RT  "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL  Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: Involved in the transcription termination process (By
CC  similarity).
CC  -1- SIMILARITY: Belongs to the nusB family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AP005279; BAB99011.1; -.
DR  HAMAP; MF_000073; -; 1.
DR  InterPro; IPR006027; NusB.
DR  Pfam; PF01029; NusB; 1.
KW  Transcription termination; Complete proteome.
SQ  SEQUENCE 227 AA; 25010 MW; F25A070B52342A5C CRC64;

Query Match          1.4%; Score 9; DB 1; Length 227;
Best Local Similarity 100.0%; Pred.No.0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  170 DSDALDSD 178
   |||||
Db  179 DSDALDSD 187

RESULT 2
RBFA OCEIH          STANDARD;          PRT;          113 AA.
ID  RBFA OCEIH
AC  Q8E0T9;
DT  10-OCT-2003 (Rel. 42, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Ribosome-binding factor A.
GN  RBFA OR OBI600.

```

OS Oceanobacillus thelyensis.
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 CC NCBI_TaxID=182710;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=HTE831 / DSM 14371 / JCM 11309;
 RX MEDLINE=22220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
 CC with 30S subunits that are part of 70S ribosomes or polysomes).
 CC Essential for efficient processing of 16S rRNA. May interact with
 CC the 5' terminal helix region of 16S rRNA (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: Belongs to the rbfA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF004598; BAC13556.1; -
 DR HAMAP; MF_00003; -; 1.
 DR InterPro; IPR000238; Rib_bind_facta.
 DR Pfam; PF02033; RBFA; 1.
 DR ProDom; PD007327; Rib_bind_facta; 1.
 DR TIGRFAMs; TIGR00082; RbfA; 1.
 DR PROSITE; PS01319; RBFA; 1.
 DR TrnA processing; Complete proteome.
 DR rRNA processing; Complete proteome.
 SQ SEQUENCE 113 AA; 12924 MW; DAC02604B386F828 CRC64;
 Query Match 1.3%; Score 8; DB 1; Length 113;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 606 EFDEALEQ 613
 DB 93 EFDEALEQ 100
 RESULT 3
 FAMB_CIOAB STANDARD; PRT; 141 AA.
 ID FAMB_CIOAB
 AC 097DA9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ((3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (EC 4.2.1.-)
 DE ((3R)-hydroxymyristoyl ACP dehydratase).
 GN FAMB OR CAC3571.
 OS Clostridium acetobutylicum.
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CC NCBI_TaxID=1488;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=1146286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -1- FUNCTION: Involved in saturated fatty acids biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

CC -1- SIMILARITY: Belongs to the thioester dehydratase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE007854; AAK81494.1; -
 DR PIR; C97338; C97338.
 DR HAMAP; MF_00406; -; 1.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 1.
 DR Lyase; Lipid synthetase; Lipid A biosynthesis; Complete proteome.
 FT ACT SITE 49 49 BY SIMILARITY.
 SQ SEQUENCE 141 AA; 15438 MW; 04928B855C2B9F5A CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 141;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 213 SLSTEQI 219
 DB 2 SLSTEQI 8
 RESULT 4
 FLAW_DESGI STANDARD; PRT; 147 AA.
 ID FLAW_DESGI
 AC 001056;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Flaxodoxin.
 DE Desulfococcus gigas.
 OS Bacteria; Proteobacteria; Deltaproteobacteria; Desulfococciales;
 CC Bacteria; Proteobacteria; Desulfococciales; Desulfococcus.
 CC Desulfococciales; Desulfococcus.
 CC NCBI_TaxID=879;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 29494 / DSM 496;
 RX MEDLINE=92329549; PubMed=1627649;
 RA Helms L.R., Swenson R.P.;
 RT "The primary structures of the flaxodoxins from two strains of
 RT Desulfococcus gigas. Cloning and nucleotide sequence of the
 RT structural genes";
 RL Biochim. Biophys. Acta 1131:325-328(1992).
 CC -1- FUNCTION: Low-potential electron donor to a number of redox
 CC enzymes.
 CC -1- COFACTOR: FMN.
 CC -1- SIMILARITY: Belongs to the flaxodoxin family.
 CC -1- SIMILARITY: Contains 1 flaxodoxin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64765; CAA46012.1; -
 DR PIR; S24310; S24310.
 DR HSSP; P00323; 2FX2.
 DR InterPro; IPR008254; Flav_nitox_synth.
 DR InterPro; IPR001094; Flaxodoxin like.
 DR InterPro; IPR001226; Flaxodoxin.
 DR Pfam; PF00258; Flaxodoxin; 1.
 DR PRINTS; PS00369; FLAVODOXIN.
 DR PROSITE; PS00201; FLAVODOXIN; 1.
 DR PROSITE; PS50902; FLAVODOXIN_LIKE; 1.

KW Electron transport; Flavoprotein; FMN.
 FT DOMAIN 4 143 FLAVODOXIN-LIKE.
 SQ SEQUENCE 147 AA; 15186 MW; 5779A72DD1395635 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 462 AKVLEES 468
 DB 23 AKVLEES 29

RESULT 5
 COX6_YEAST
 ID COX6_YEAST STANDARD; PRT; 148 AA.
 AC P00427;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide VI, mitochondrial precursor
 (EC 1.9.3.1).
 GN COX6 OR YHR051W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85080033; PubMed=6210289;
 RA Wright R.M., Ko C., Cuskey M.G., Poyton R.O.;
 RT "Isolation and sequence of the structural gene for cytochrome c
 RT oxidase subunit VI from Saccharomyces cerevisiae.";
 RL J. Biol. Chem. 259:15401-15407(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=6288c / AB972;
 MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 Du Z., Favello A., Fulton L., Gattung S., Geisler C., Kirsten J.,
 Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
 Latteille P., Louis E.J., Macri C., Marzola E., Meneses S., Mouser L.,
 Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 RN [3]
 RP SEQUENCE OF 1-23 FROM N.A.
 RX MEDLINE=89160242; PubMed=2537949;
 RA Wright R.M., Rosenzweig B., Poyton R.O.;
 RT "Organization and expression of the COX6 genetic locus in
 RT Saccharomyces cerevisiae: multiple mRNAs with different 3' termini
 RT are transcribed from COX6 and regulated differentially.";
 RL Nucleic Acids Res. 17:1103-1120(1989).
 RN [4]
 RP SEQUENCE OF 41-148.
 RX MEDLINE=83030850; PubMed=6230493;
 RA Gregor I., Tsugita A.;
 RT "The amino acid sequence of cytochrome c oxidase subunit VI from
 RT Saccharomyces cerevisiae.";
 RL J. Biol. Chem. 257:11081-11087(1982).
 CC -1- FUNCTION: This is the heme A-containing chain of cytochrome c
 CC oxidase, the terminal oxidase in mitochondrial electron transport.
 CC -1- CATALYTIC ACTIVITY: 4 ferricytochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- SUBUNIT: Composed of at least 11 subunits.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC -1- SIMILARITY: Belongs to the cytochrome c oxidase Va family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M10138; AAA66900.1; -
 CC EMBL; U00062; AAB68899.1; -
 CC EMBL; X14452; CAA32622.1; -
 CC PIR; A22853; OTBYC.
 CC Germonline; 139368; -

DR SGD; S0001093; COX6.
 DR GO; GO:0005751; C:respiratory chain complex IV (sensu Eukarya); IPI.
 DR GO; GO:0004129; F:cytochrome-c oxidase activity; IDA.
 DR InterPro; IPR003204; Cyt c_ox5A.

DR Pfam; PF02284; COX5A; 1.
 DR Oxidoreductase; Heme; Mitochondrion; Inner membrane; Transit peptide.
 FT TRANSIT 1 40 MITOCHONDRION.
 FT CHAIN 41 148 CYTOCHROME C OXIDASE POLYPEPTIDE VI.
 SQ SEQUENCE 148 AA; 17341 MW; 91F2AF286BD2FB1A CRC64;

Query Match 1.1%; Score 7; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 604 EKEPDEA 610
 DB 57 EKEPDEA 63

RESULT 6
 G6PD_SARBU
 ID G6PD_SARBU STANDARD; PRT; 153 AA.
 AC Q25537;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD)
 (Zwischenferment) (Fragment).
 GN 2W.
 OS Sarcophaga bullata (grey flesh fly) (Neobellieria bullata).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Sarcophagidae; Sarcophaga.
 OX NCBI_TaxID=7385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Soto-Adames F.N., Robertson H.M., Berlocher S.H.;
 RT "Phylogenetic utility of partial DNA sequences of G6PDH at different
 RT taxonomic levels in Hexapoda with emphasis on Diptera.";
 RL Ann. Entomol. Soc. Am. 87:723-736(1994).
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
 CC 1,5-lactone 6-phosphate + NADPH.
 CC -1- PATHWAY: Pentose phosphate pathway; first step.
 CC -1- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U09037; AAB02781.1; -
 CC HSSP; P11411; IDPG.
 DR InterPro; IPR001282; G6PD.
 DR Pfam; PF00479; G6PD; 1.
 DR PRINTS; PR00079; G6PDHGNASE.
 DR ProDom; PD001129; G6PD; 1.
 DR PROSITE; PS00069; G6P DEHYDROGENASE; PARTIAL.
 KW Oxidoreductase; NADP; Glucose metabolism.

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0273 protein TM0370.
GN TM0370.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
NCBI_TaxID=2336;
RX NCB1_TaxID=2336;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MS98 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316, PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L.M., Utterback T.R., Malek J.A., Linher K.D., Garrett M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of *Thermotoga maritima*.;
RL Nature 399:323-329(1999).
RT
RL Nature 399:323-329(1999).
CC -1- SIMILARITY: Belongs to the UPF0273 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sb.ch).
CC
CC -----
CC EMBL; AE001717; AAC5457.1; ALT_INT.
DR PIR; H22384; H7384.
DR TIGR; TM0370; -.
DR HAMAP; MF_01076; -; 1.
DR Hypothetical protein; ATP-binding; Complete proteome.
KW NP_BIND 30 ATP (POTENTIAL).
FT MF_BIND 30
SQ SEQUENCE 242 AA; 26813 MW; 396431D4D5E48837 CRC64;
QY 262 RELAEVL 268 1.1%; Score 7; DB 1; Length 242;
Db 114 RELAEVL 120 Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
RESULT 9
NT3_CHICK PRT; 257 AA.
ID NT3_CHICK STANDARD;
AC P25433;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE Nerve growth factor 2) (NGF-2).
DE NT3.
GN Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCB1_TaxID=9031;
OX (1)
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=930912381, PubMed=1457809;
RX Maisongier P., Bellucio L., Conover J.C., Yancopoulos G.D.;
RA "Gene sequence of chicken BDNF and NT-3.";
RT DNA Seq. 3:49-54(1992).
RN
RP SEQUENCE OF 194-236 FROM N.A.
RX MEDLINE=91222573; PubMed=2025430;
RX Haldoeok F., Ibanez C.F., Persson H.;

```

RT "Evolutionary studies of the nerve growth factor family reveal a
RT novel member abundantly expressed in Xenopus ovary."
RL Neuron 6:845-858(1991).
CC -1- FUNCTION: Seems to promotes the survival of visceral and
CC proprioceptive sensory neurons.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63378; AAA68880.1; -.
DR F1R; I50400; I50400.
DR HSSP; P20783; 1B8K.
DR InterPro; IPR002072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PR00268; NGF; 1.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
DR Growth factor; Signal_1.
DR SIGNAL 1 16
DR PROPEP 17 138
DR CHAIN 139 257
DR DISULFID 152 217
DR DISULFID 195 246
DR DISULFID 205 248
DR CARBOHYD 131 131
SQ SEQUENCE 257 AA; 29701 MW; E043BA2A005C1E7 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
Db 45 KNTLSKQ 51

RESULT 10
NT3_FELCA STANDARD; PRT; 257 AA.
ID NT3_FELCA
AC 09TST2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE (Nerve growth factor 2) (NGF-2).
GN NTF3.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxId=9685;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20211727; PubMed=10745216;
RA Lein E.S., Hohn A., Shatz C.J.;
RT "Dynamic regulation of BDNF and NT-3 expression during visual system
RT development."
RL J. Comp. Neurol. 420:1-18(2000).
CC -1- FUNCTION: Seems to promotes the survival of visceral and
CC proprioceptive sensory neurons (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF192538; AAF03424.1; -.
DR HSSP; P20783; 1B8K.
DR InterPro; IPR002072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PR00268; NGF; 1.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
DR Growth factor; Signal_1.
DR SIGNAL 1 16
DR PROPEP 17 138
DR CHAIN 139 257
DR DISULFID 152 217
DR DISULFID 195 246
DR DISULFID 205 248
DR CARBOHYD 131 131
SQ SEQUENCE 257 AA; 29403 MW; E53F7B59C5113E4 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
Db 45 KNTLSKQ 51

RESULT 11
NT3_HUMAN STANDARD; PRT; 257 AA.
ID NT3_HUMAN
AC P20783;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE (Nerve growth factor 2) (NGF-2).
GN NTF3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90262727; PubMed=2344409;
RA Rosenthal A., Goeddel D.V., Nguyen T., Lewis M., Shih A.,
RA Iarame G.R., Nikolic K., Winslow J.W.;
RT "Primary structure and biological activity of a novel human
RT neurotrophic factor."
RL Neuron 4:767-773(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91045937; PubMed=2236018;
RA Jones K.R., Reichardt L.F.;
RT "Molecular cloning of a human gene that is a member of the nerve
RT growth factor family."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8060-8064(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90306351; PubMed=2365067;
RA Kaisho Y., Yoshimura K., Nakahama K.;
RT "Cloning and expression of a cDNA encoding a novel human neurotrophic
RT factor."
RL FEBS Lett. 266:187-191(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91365361; PubMed=1889806;

```

RA Maisonnier P.C., le Beau M.M., Espinosa R. III, Ip N.Y.,
RA Belluscio L., de la Monte S.M., Squinto S., Purth M.E.,
RA Yancopoulos G.D.;
RT "Human and rat brain-derived neurotrophic factor and neurotrophin-3:
RT gene structures, distributions, and chromosomal localizations.";
RL Genomics 10:558-568(1991).
RN [5]
RP SEQUENCE OF 194-236 FROM N.A.
RC TISSUE=Leukocyte; PubMed=2025430;
RX MEDLINE=91222573; PubMed=2025430;
RA Hallboeck F., Ibanez C.F., Persson H.;
RT "Evolutionary studies of the nerve growth factor family reveal a
RT novel member abundantly expressed in Xenopus ovary.";
RL Neuron 6:845-858(1991).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=95217877; PubMed=7703225;
RA Robinson R.C., Radziejewski C., Stuart D.I., Jones E.Y.;
RT "Structure of the brain-derived neurotrophic factor/neurotrophin 3
RT heterodimer.";
RL Biochemistry 34:4139-4146(1995).
RN [7]
RP VARIANT GLU-76.
RX MEDLINE=95251647; PubMed=7733919;
RA Hattori M., Nanko S.;
RT "Association of neurotrophin-3 gene variant with severe forms of
RT schizophrenia.";
RL Biochem. Biophys. Res. Commun. 209:513-518(1995).
RN [8]
RP VARIANT GLU-76.
RX MEDLINE=96253892; PubMed=8925252;
RA Arinami T., Takekoshi K., Itokawa M., Hamaguchi H., Toru M.;
RT "Failure to find associations of the CA repeat polymorphism in the
RT first intron and the Gly-63/Glu-63 polymorphism of the neurotrophin-3
RT gene with schizophrenia.";
RL Psychiatr. Genet. 6:13-15(1996).
RN [9]
RP FUNCTION: Seems to promotes the survival of visceral and
RT proprioceptive sensory neurons.
RL SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.
CC -1- POLYMORPHISM: Variant Glu-76 (frequently reported as Glu-63) was
CC thought to be associated with severe forms of schizophrenia. This
CC does not seem to be the case.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53653; CAA37703.1; -
CC EMBL; M37763; AA595953.1; -
CC EMBL; M61180; AA63231.1; -
CC PIR; A36208; C40304.
CC PDB; 1BND; 04-APR-96.
CC PDB; 1B8K; 09-FEB-99.
CC PDB; 1NT3; 16-JUN-00.
CC Gene; HGNC:8023; NTF3.
CC MIM; 162660; -
CC GO; GO:0005102; F:receptor binding; TAS.
CC GO; GO:0007267; P:cell-cell signaling; TAS.
CC GO; GO:0007399; P:neurogenesis; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR002072; NGF.
CC Pfam; PF00243; NGF; 1.
CC PRINTS; PR00268; NGF.
CC PRODOM; PD002052; NGF; 1.
CC SMART; SM00140; NGF; 1.
CC PROSITE; PS00248; NGF_1; 1.
CC PROSITE; PS50270; NGF_2; 1.
CC DR

KW Growth factor; Signal; Polymorphism; 3d-structure.
FT SIGNAL 1 16
FT PROPEP 17 138
FT CHAIN 139 257
FT DISULFID 152 217
FT DISULFID 195 246
FT DISULFID 205 248
FT CARBOHYD 131 131
FT VARIANT 76 76
FT STRAND 148 150
FT STRAND 154 159
FT STRAND 164 167
FT STRAND 168 169
FT TURN 172 175
FT STRAND 178 179
FT TURN 181 183
FT STRAND 186 187
FT STRAND 190 195
FT TURN 206 207
FT STRAND 208 208
FT TURN 210 211
FT TURN 214 230
FT STRAND 231 232
FT TURN 233 251
FT SEQUENCE 257 AA; 29354 MW; 39A5B3B28E25E03 CRC64;
SQ
Query Match 1.1%; Score 7; DB 1; Length 257;
Beat Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 452 KNKLSKQ 458
DB 45 KNKLSKQ 51
RESULT 12
ID NT3 MOUSE STANDARD; PRT; 258 AA.
AC P20181;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE (Nerve growth factor 2) (NGF-2).
GN NTF3 OR NTF-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90190865; PubMed=2314473;
RA Hohn A., Leibrock J., Bailey K., Barde Y.-A.;
RT "Identification and characterization of a novel member of the nerve
RT growth factor/brain-derived neurotrophic factor family.";
RL Nature 344:339-341(1990).
CC -1- FUNCTION: Seems to promotes the survival of visceral and
CC proprioceptive sensory neurons.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53257; CAA37348.1; -
CC PIR; S09155; S09155.
CC DR

```

DR HSSP; P20783; 1B8K.
DR MGD; MGI:97380; Nef3.
DR GO; GO:000515; F:protein binding; IMP.
DR GO; GO:0007420; P:brain development; IMP.
DR GO; GO:0008544; P:epidermal differentiation; IMP.
DR GO; GO:0007403; P:glial cell fate determination; IMP.
DR GO; GO:0007422; P:peripheral nervous system development; IMP.
DR GO; GO:0045944; P:positive regulation of transcription from P. .; IDA.
DR InterPro; IPR002072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PR00268; NGF.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
KW Growth factor; Signal.
FT SIGNAL 1 16
FT PROPEP 17 139
FT CHAIN 140 258
FT DISULFID 153 218
FT DISULFID 196 247
FT DISULFID 206 249
FT CARBOHYD 131 133
SQ SEQUENCE 258 AA; 29587 MW; 7180D064E8AE6042 CRC64; (POTENTIAL).

Query Match 1.1%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
DB 45 KNTLSKQ 51

RESULT 13
NT3_RAT STANDARD; PRT; 258 AA.
AC P18280;
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE (Nerve growth factor 2) (NGF-2).
GN NTF3 OR NTF-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90319130; PubMed=2164684;
RA Ernster P., Ibanez C.F., Ebendal T., Olson L., Persson H.;
RT "Molecular cloning and neurotrophic activities of a protein with
RT structural similarities to nerve growth factor: developmental and
RT topographical expression in the brain.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:5454-5458(1990).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=90208301; PubMed=2321006;
RA Malsenpleire P.C., Belluscio L., Squinto S., Ip N.Y., Furth M.E.,
RA Linday R.M., Yancopoulos G.D.;
RT "Neurotrophin-3: a neurotrophic factor related to NGF and BDNF.";
RT Science 247:1446-1451(1990).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=91365361; PubMed=1889806;
RA Malsenpleire P.C., le Beau M.M., Espinosa R. III, Ip N.Y.,
RA Belluscio L., de la Monte S.M., Squinto S., Furth M.E.,
RA Yancopoulos G.D.;
RT "Human and rat brain-derived neurotrophic factor and neurotrophin-3:
RT gene structures, distributions, and chromosomal localizations.";
RT Genomics 10:558-568(1991).
[4]

```

```

RP SEQUENCE OF 195-237 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=91222573; PubMed=2025430;
RA Halboeck F., Ibanez C.F., Persson H.;
RT "Evolutionary studies of the nerve growth factor family reveal a
RT novel member abundantly expressed in Xenopus ovary.";
RL Neuron 6:845-858(1991).
CC -1- FUNCTION: Seems to promotes the survival of visceral and
CC proprioceptive sensory neurons.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M36643; AAA41313.1; -
DR EMBL; M39668; AAA41727.1; -
DR EMBL; M61179; AAA63497.1; -
DR PIR; A35781; A35781.
DR HSSP; P20783; 1B8K.
DR InterPro; IPR002072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PR00268; NGF.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
KW Growth factor; Signal.
FT SIGNAL 1 16
FT PROPEP 17 139
FT CHAIN 140 258
FT DISULFID 153 218
FT DISULFID 196 247
FT DISULFID 206 249
FT CARBOHYD 131 131
SQ SEQUENCE 258 AA; 29644 MW; 74D557CF8518A1CE CRC64; (POTENTIAL).

Query Match 1.1%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
DB 45 KNTLSKQ 51

RESULT 14
CIT2_MOUSE STANDARD; PRT; 269 AA.
AC Q35740; Q35741; Q35742; Q35743; Q55198;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cbp/p300-interacting transactivator 2 (MSG-related protein 1) (MRG1
DE protein).
GN CITED2 OR MRG1 OR MSG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC STRAIN=C57BL/6 X DBA;
RX MEDLINE=98202510; PubMed=9533950;
RA Dunwoodie S.L., Rodriguez T.A., Beddington R.S.P.;
RA "Mg1 and Mtg1, founding members of a gene family, show distinct
RT patterns of gene expression during mouse embryogenesis.";

```

Mech. Dev. 72:27-40(1998).
RN SEQUENCE FROM N.A. (ISOFORM 3).
RP MEDLINE=98094278; PubMed=9434189;
RX Shioda T., Fenner M.H., Isselbacher K.J.;
RA "MSG1 and its related protein MRG1 share a transcription activating
RT domain";
RL Gene 204:1235-241(1997).
CC -1- FUNCTION: Interferes with the binding of transcription factors
CC HIF-1a and STAT2 to the p300/CBP CH1 region (By similarity).
CC -1- SUBUNIT: Binds to the p300/CBP CH1 region (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=O35740-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O35740-2; Sequence=VSP_001090;
CC Name=3;
CC IsoId=O35740-3; Sequence=VSP_001091;
CC Name=4;
CC IsoId=O35740-4; Sequence=VSP_001092;
CC -1- SIMILARITY: Belongs to the CITED family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y15163; CAAT5432.1; -
CC EMBL; Y15163; CAAT5433.1; -
CC EMBL; Y15163; CAAT5434.1; -
CC EMBL; Y15163; CAAT5435.1; -
CC EMBL; U86445; AAC39945.1; -
CC MGD; MGI:1306784; Cited2.
CC GO; GO:0007417; P:central nervous system development; IMP.
CC DR GO; GO:0007417; P:central nervous system development; IMP.
CC DR InterPro; IPR007576; CITED.
CC DR Pfam; PF04487; CITED; 1.
CC DR Transcription regulation; Nuclear protein; Alternative splicing.
CC KM DOMAIN 21 57 HIS-RICH.
CC FT DOMAIN 162 197 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 218 257 Missing (in isoform 2).
CC FT VARSPLIC 138 158 /FridaVSP_001090.
CC FT VARSPLIC 159 213 Missing (in isoform 3).
CC FT VARSPLIC 203 213 Missing (in isoform 4).
CC FT VARSPLIC 203 213 /FridaVSP_001092.
CC SQ SEQUENCE 269 AA; 28382 MW; AFACDDSD7902448 CRC64;
Query Match 1.1%; Score 7; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 378 EMBLDR1 384
DB 233 EMBLDR1 239
RESULT 15
CIT2_HUMAN STANDARD; PRT; 270 AA.
AC Q99967; O95426;
ID 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cbp/p300-interacting transactivator 2 (MSG-related protein 1) (MRG1
DE protein) (p35e-f).
GN CITED2 OR MRG1.
OS Homo sapiens (human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RP MEDLINE=97057236; PubMed=8901575;
RX Shioda T., Fenner M.H., Isselbacher K.J.;
RA "msg1, a novel melanocyte-specific gene, encodes a nuclear protein
RT and is associated with pigmentation";
RL Proc. Natl. Acad. Sci. U.S.A. 93:12298-12303(1996).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP MEDLINE=20021768; PubMed=10552932;
RX Leung M.K., Jones T., Michels C.L., Livingston D.M., Bhattacharya S.;
RA "Molecular cloning and chromosomal localization of the human CITED2
RT gene encoding p35e/f/Mrg1";
RL Genomics 61:307-313(1999).
RN [3]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP MEDLINE=99104162; PubMed=9887100;
RX Bhattacharya S., Michels C.M., Leung M.K., Arany Z.P., Kung A.L.,
RA Bhattacharya S., Michels C.M., Leung M.K., Arany Z.P., Kung A.L.,
RT "Functional role of p35e/f, a novel p300/CBP binding protein, during
RL transactivation by HIF-1";
RN Genes Dev. 13:64-75(1999).
RN [4]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE-Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Straube R.L., Rollins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Bietow K.H., Schaefer C.F., Hsieh F.,
RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkin R.F., Jordan A., Farmer A.A., Rubin G.M., Hong L.,
RA Datchenko L., Marusina K., Carninci P., Casavant T.L., Scheetz T.E.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Carninci P., Prange C.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Abramson R.D., Miliady S.J.,
RA Rana S.S., Loggiano N.A., Peters G.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McMan P.J., McKernan K.J., Gay L.J., Hulyk S.W.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman W., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Smalhe D.E.,
RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences";
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Interferes with the binding of transcription factors
CC HIF-1a and STAT2 to the p300/CBP.
CC -1- SUBUNIT: Binds to the p300/CBP CH1 region.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O99967-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O99967-2; Sequence=VSP_001089;
CC -1- INDUCTION: By hypoxia and deferoxamine.
CC -1- SIMILARITY: Belongs to the CITED family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U65093; AAC5114.1; -
CC EMBL; AF129290; AAF01263.1; -
CC EMBL; AF129290; AAF01264.1; -

DR EMBL; AF109161; AAD10055.1; -.
 DR EMBL; BC004377; AA04377.1; -.
 DR Genew; HGNC:1987; CITED2.
 DR MIM: 602937; -.
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0005515; F:protein binding; NAS.
 DR GO; GO:0003700; F:transcription factor activity; TAS.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
 DR InterPro; IPR007576; CITED.
 DR Pfam; PF04487; CITED; 1.
 KW Transcription regulation; Nuclear protein; Alternative splicing.
 FT DOMAIN 21 57 HIS-RICH.
 FT DOMAIN 162 199 GLY-RICH.
 FT DOMAIN 219 258 ASP/GLY-RICH (ACIDIC).
 FT VARSPLIC 159 215 Missing (in isoform 2).
 FT /FTId-VSP_001089
 SQ SEQUENCE 270 AA: 28497 MW: 45DE3A5E2B4C472 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 270;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 EMGLDRI 384
 |||||
 Db 234 EMGLDRI 240

Search completed: April 6, 2004, 19:42:29
 Job time : 20 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:39:22 ; Search time 46 Seconds
(without alignments)
4286.932 Million cell updates/sec

Title: US-10-030-829-3
Perfect score: 625
Sequence: 1 MESSRAGPMSEKKNVQGGYRP.....EFDEALEQLMYKHGLHNEDD 625

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	625	100.0	625	10 Q9LDX1	Q9Ldx1 arabidopsis
2	13	2.1	170	10 Q7XYE6	Q7xye6 triticum ae
3	10	1.6	304	10 Q7XY17	Q7xy17 triticum ae
4	9	1.4	272	5 Q7Z144	Q7z144 caenorhabdi
5	9	1.4	321	16 Q8F052	Q8f052 corynebacte
6	9	1.4	382	16 Q98B79	Q98b79 rhizobium l
7	8	1.3	212	16 Q8G635	Q8g635 bifidobacte
8	8	1.3	273	16 Q7WEP5	Q7wep5 bordetella
9	8	1.3	273	16 Q7W3C6	Q7w3c6 bordetella
10	8	1.3	273	16 Q7VU31	Q7vu31 bordetella
11	8	1.3	296	10 Q9FNU1	Q9fnu1 oryza sativ
12	8	1.3	408	17 Q8Z2M7	Q8z2m7 pyrobaculum
13	8	1.3	448	6 Q9XSE3	Q9xse3 equus caball
14	8	1.3	516	10 Q8LILN5	Q8liln5 oryza sativ
15	8	1.3	525	16 Q8R9U3	Q8r9u3 thermoaer
16	8	1.3	586	5 Q18097	Q18097 caenorhabdi

17	8	1.3	616	5 Q96148	Q96148 plasmodium
18	8	1.3	780	5 Q811Z2	Q811z2 plasmodium
19	8	1.3	855	10 Q8S2B7	Q8s2b7 oryza sativ
20	8	1.3	916	16 Q7WGI6	Q7wgi6 bordetella
21	8	1.3	916	16 Q7W521	Q7w521 bordetella
22	8	1.3	916	16 Q7VVY4	Q7vvy4 bordetella
23	7	1.1	69	17 Q9HPS2	Q9hps2 halobacteri
24	7	1.1	91	5 Q8T9N2	Q8t9n2 plasmodium
25	7	1.1	92	16 Q8R5Q2	Q8r5q2 thermoaer
26	7	1.1	92	16 Q8DC09	Q8dc09 vibrio vuln
27	7	1.1	93	4 Q9NS17	Q9ns17 homo sapien
28	7	1.1	94	16 Q8TA19	Q8ta19 xyella tas
29	7	1.1	101	17 Q9HSX0	Q9hsx0 halobacteri
30	7	1.1	108	2 Q8L342	Q8l342 vibrio chol
31	7	1.1	114	11 Q9D2T3	Q9d2t3 mus musculu
32	7	1.1	121	4 Q8N7K2	Q8n7k2 homo sapien
33	7	1.1	126	10 Q9MAD2	Q9mad2 arabidopsis
34	7	1.1	129	11 Q9WU16	Q9wu16 mesocricetu
35	7	1.1	135	16 Q67676	Q67676 aquifex aeo
36	7	1.1	139	6 Q8SPT8	Q8spt8 macaca mula
37	7	1.1	146	5 Q8T9Z8	Q8t9z8 plasmodium
38	7	1.1	147	13 Q7SZP9	Q7szp9 fuqu rubrip
39	7	1.1	147	16 Q8UCU6	Q8ucu6 agrobacteri
40	7	1.1	148	11 Q8BP19	Q8bp19 mus musculu
41	7	1.1	149	16 Q07614	Q07614 bacillus su
42	7	1.1	152	5 Q92T22	Q92t22 rhizobium m
43	7	1.1	157	5 Q967Z3	Q967z3 plasmodium
44	7	1.1	159	10 Q84T15	Q84t15 phaseolus a
45	7	1.1	163	5 Q7YV99	Q7yv99 trypanosoma

ALIGNMENTS

RESULT 1	ID	Q9LDX1	PRELIMINARY;	PRT;	625 AA.
AC	Q9LDX1				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)			
DE	EMBL CAB62356.1 (SGS3)	(Hypothetical protein).			
GN	AT5G23570.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	euroside II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=COLIMBIA;				
RX	MEDLINE=20181125; PubMed=10718197;				
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,				
RA	Tabata S.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence				
RT	features of the regions of 3,076,755 bp covered by sixty Pl and TAC				
RT	clones.";				
RL	DNA Res. 7:31-63(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=20306668; PubMed=10850495;				
RA	Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,				
RA	Morel U.-B., Jouette D., Lacombe A.-M., Nikic S., Picault N.,				
RA	Remoue K., Sanjal M., Vo T.-A., Vaucheret H.;				
RT	"Arabidopsis SGS2 and SGS3 Genes are Required for Posttranscriptional				
RT	Gene Silencing and Natural Virus Resistance.";				
RL	Cell 101:533-542(2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Yamada K., Chan M.-M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,				
RA	Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,				
RA	Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,				

RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Arabidopsis Full Length cDNA Clones";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Heuan V.W., Lee J.M.,
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB025563; BAA97244.1; -;
 DR EMBL; AF239719; AAF73960.1; -;
 DR EMBL; BT002944; AAO22757.1; -;
 DR EMBL; BT004380; AAO42374.1; -;
 DR InterPro; IPR005380; XS.
 DR InterPro; IPR005381; zF.
 DR Pfam; PF03468; XS; 1.
 DR Pfam; PF03470; zF-XS; 1.
 DR Hypoetical protein;
 KW SEQUENCE 625 AA; 71971 MW; 456E2A1396706A96 CRC64;
 SQ
 Query Match 100.0%; Score 625; DB 10; Length 625;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MSSRAGPMSKKNVGGYRPEYVQLAGTRLASQDDGEMEVISKKNKPKNTSG 60
 1 MSSRAGPMSKKNVGGYRPEYVQLAGTRLASQDDGEMEVISKKNKPKNTSG 60
 61 KTWVSGNSNPRAWGQQGGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120
 61 KTWVSGNSNPRAWGQQGGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120
 61 KTWVSGNSNPRAWGQQGGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120
 121 NFVAPPVSRPPLLEGGMNQARGSAOHTAVOEPFVDEDDVDNASEEENDSDALDSDDD 180
 121 NFVAPPVSRPPLLEGGMNQARGSAOHTAVOEPFVDEDDVDNASEEENDSDALDSDDD 180
 121 NFVAPPVSRPPLLEGGMNQARGSAOHTAVOEPFVDEDDVDNASEEENDSDALDSDDD 180
 181 IASDDVDSDVSGSHSRKQNKFFKFGSLDLSISIOINERQWHPACONGEGALDW 240
 181 IASDDVDSDVSGSHSRKQNKFFKFGSLDLSISIOINERQWHPACONGEGALDW 240
 181 IASDDVDSDVSGSHSRKQNKFFKFGSLDLSISIOINERQWHPACONGEGALDW 240
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLQMRGASVPCGELTYGQWKGIGDEKDY 300
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLQMRGASVPCGELTYGQWKGIGDEKDY 300
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLQMRGASVPCGELTYGQWKGIGDEKDY 300
 301 IVPMPVITINTRLDKDNDKWLGMGNQELLEFDYELRLARHSGPGCHGMSVLMFE 360
 301 IVPMPVITINTRLDKDNDKWLGMGNQELLEFDYELRLARHSGPGCHGMSVLMFE 360
 301 IVPMPVITINTRLDKDNDKWLGMGNQELLEFDYELRLARHSGPGCHGMSVLMFE 360
 361 SSATGYEABERLHLEMLGDLRIAWGQKSMFSGVRYGLAKDKDLDIFNGSQK 420
 361 SSATGYEABERLHLEMLGDLRIAWGQKSMFSGVRYGLAKDKDLDIFNGSQK 420
 361 SSATGYEABERLHLEMLGDLRIAWGQKSMFSGVRYGLAKDKDLDIFNGSQK 420
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKQNHAVYLSLEISSEKLRRA 480
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKQNHAVYLSLEISSEKLRRA 480
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKQNHAVYLSLEISSEKLRRA 480
 481 EDNRIVRQTYQEHQNEEDDADRPFMSIKQIHERDAKEMFEMLQOQERAKVYQ 540
 481 EDNRIVRQTYQEHQNEEDDADRPFMSIKQIHERDAKEMFEMLQOQERAKVYQ 540
 481 EDNRIVRQTYQEHQNEEDDADRPFMSIKQIHERDAKEMFEMLQOQERAKVYQ 540
 541 QOONINPSSNDCCRAAEVSSFIEFOKEMEFEVEREMLIKQDEKMDMKRHHET 600
 541 QOONINPSSNDCCRAAEVSSFIEFOKEMEFEVEREMLIKQDEKMDMKRHHET 600
 541 QOONINPSSNDCCRAAEVSSFIEFOKEMEFEVEREMLIKQDEKMDMKRHHET 600
 601 FDLKEFDEALBQMLYKHLNEDD 625

DB 601 FDLKEFDEALBQMLYKHLNEDD 625
 RESULT 2
 ID 07XYE6 PRELIMINARY; PRT; 170 AA.
 AC 07XYE6;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE H0TR (Fragment).
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. PH 82-2-2;
 RA Li J.R., Wang F., Li O.Z., Zhang X.S.;
 RT "Gene Isolation and expression of a new Zn-finger";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF469493; AAP80610.1; -;
 DR NON TER 1
 FT SEQUENCE 170 AA; 19980 MW; 0AFD21F1778B3068 CRC64;
 SQ
 Query Match 100.0%; Score 13; DB 10; Length 170;
 Best Local Similarity 100.0%; Pred. No. 4,9e-05;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 322 WLMGNQELLEYP 334
 1 WLMGNQELLEYP 13
 RESULT 3
 ID 07XY17 PRELIMINARY; PRT; 304 AA.
 AC 07XY17;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Emt1.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. PH 82-2-2;
 RA Zhao X., Li O., Zhang X.;
 RT "Isolation and expression of a new kind of gene involve in
 embryoogenesis in Triticum aestivum L.";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF542974; AAP80862.1; -;
 FT SEQUENCE 304 AA; 35958 MW; 89980215A9584208 CRC64;
 SQ
 Query Match 1.6%; Score 10; DB 10; Length 304;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 325 MGNQELLEYP 334
 1 MGNQELLEYP 10
 RESULT 4
 ID 07Z144 PRELIMINARY; PRT; 272 AA.
 AC 07Z144;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)

```

DT 01-OCT-2003 (TEMBLrel. 25, last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, last annotation update)
DE Hypothetical protein ZK546.17.
GN ZK546.17.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Hallsworth K.;
RT "The sequence of C. elegans cosmid ZK546";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U9380; AAP68926.1; -
KW Hypothetical protein.
SQ SEQUENCE 272 AA; 31290 MW; CB99BA62BD7DEB84 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 5; Length 272;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 LEIMSEKLR 477
DB 102 LEIMSEKLR 110

RESULT 5
Q8FO52 PRELIMINARY; PRT; 321 AA.
AC Q8FO52;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE Putative glucose dehydrogenase-B.
GN CE1282.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005218; BAC18092.1; -
KW Complete proteome.
SQ SEQUENCE 321 AA; 34379 MW; C3D9C980AFF3C080 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 321;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 479 TAEDNRIVR 487
DB 75 TAEDNRIVR 83

RESULT 6
Q98BT9 PRELIMINARY; PRT; 382 AA.
AC Q98BT9;
DT 01-OCT-2001 (TEMBLrel. 18, Created)
DT 01-OCT-2001 (TEMBLrel. 18, last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, last annotation update)
DE Acyl-CoA dehydrogenase.
GN M15430.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003006; BAB51883.1; -
DR GO; GO:0003995; F:acyl-CoA dehydrogenase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006090; Acyl-CoA_dh_C.
DR InterPro; IPR006091; Acyl-CoA_dh_M.
DR Pfam; PF02770; Acyl-CoA_dh_M; 1.
KW Complete proteome.
SQ SEQUENCE 382 AA; 41252 MW; DB2B4A400E1864C7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 382;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 HRELAEVLE 269
DB 16 HRELAEVLE 24

RESULT 7
Q8G635 PRELIMINARY; PRT; 212 AA.
AC Q8G635;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE Hypothetical protein.
GN BL0815.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteriia; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Scheil M.A., Karmaliantzou M., Snel B., Vilianova D., Berger B.,
RA Pessi G., Zwaan M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
to the human gastrointestinal tract."

```

```

RL  Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
RM  EMBL; AE014703; LAN24629.1; -.
KW  Hypothetical protein. Complete proteome.
SQ  SEQUENCE 212 AA; 25137 MW; EDA495CDB8D8B53D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 212;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 ASDDYDSD 189
DB 195 ASDDYDSD 202

RESULT 8
Q7WEP5 PRELIMINARY; PRT; 273 AA.
ID Q7WEP5
AC Q7WEP5;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN B45589.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-589;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebailia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640451; CAE34951.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29859 MW; EA0423C75C3AC225 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 9
Q7W3C6 PRELIMINARY; PRT; 273 AA.
ID Q7W3C6
AC Q7W3C6;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BPA119.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;

```

```

RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebailia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640435; CAE39398.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29902 MW; 692511C67C253C3A CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 10
Q7VU31 PRELIMINARY; PRT; 273 AA.
ID Q7VU31
AC Q7VU31;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BP3304.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebailia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43569.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29859 MW; EA0423C75C3AC225 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 11
Q9FN11 PRELIMINARY; PRT; 296 AA.
ID Q9FN11

```

```

AC Q9FNUL;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 3615.5.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare; TRANSPOSON=gaigan;
RX MEDLINE=21140318; PubMed=11244114;
RA Dubcovsky J., Ramakrishna W., SanMiguel P.J., Busso C.S., Yan L.,
RA Shiloff B.A., Bennett J.L.;
RT "Comparative sequence analysis of colinear barley and rice bacterial
RT artificial chromosomes."
RL Plant Physiol. 125:1342-1353(2001).
DR EMBL; AY013245; AAC45493.1; -.
DR Gramene; Q9FNUL; -.
SQ SEQUENCE 296 AA; 3326 MW; CC472BACED677608 CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 10; Length 296;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 SDALDSD 178
Db 254 SDALDSD 261

RESULT 12
Q8Z2M7 PRELIMINARY; PRT; 408 AA.
ID Q8Z2M7
AC Q8Z2M7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PAE0172.
GN PAE0172.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Filiz-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009754; AL62612.1; -.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR SMART; SM00382; AAA, 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 408 AA; 46291 MW; 02A24E5A2523386F CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 17; Length 408;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 HRELAEVL 268
Db 86 HRELAEVL 93

RESULT 13
Q9XS63 PRELIMINARY; PRT; 448 AA.
ID Q9XS63

```

```

AC Q9XS63;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chromogranin A.
GN CGA.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal medulla;
RX MEDLINE=20492714; PubMed=11039590;
RA Sato F., Hasegawa T., Katayama Y., Iwanaga T., Yanaihara N., Kanno T.,
RA Ichida N.;
RT "Equus caballus CGA mRNA for chromogranin A, complete cds."
RL J. Vet. Med. Sci. 62:953-959(2000).
DR EMBL; AB025570; BAA76748.1; -.
DR InterPro; IPR001819; Chromogranin_AB.
DR InterPro; IPR001990; Granin.
DR Pfam; PF01271; Granin; 1.
DR PRINTS; PR00659; CHROMOGRANIN.
DR PROSITE; PS00422; GRANINS_1; 1.
DR PROSITE; PS00423; GRANINS_2; 1.
SQ SEQUENCE 448 AA; 49861 MW; EC2D6418F5BA5274 CRC64;

Query Match
Best Local Similarity 1.3%; Score 8; DB 6; Length 448;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 ELAEVLEK 270
Db 102 ELAEVLEK 109

RESULT 14
Q8L1N5 PRELIMINARY; PRT; 516 AA.
ID Q8L1N5
AC Q8L1N5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein OJ112 G08.17.
GN OSJNBA0032E21.01 OR OJ1112 G08.17.
OS Oryza sativa (Japanese cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Eastman A.P., Smith S.C., Bertin N., Liang C., Najjar F.Z., Pratt L.H.,
RA Cordouanier-Eratt M.-M.;
RT "Uncloned."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Telcitr T., Kim M.M., Bera J.J., Jin S.S.,
RA Padrosh D.W., Tallon L.J., Koo H., Ziemann V., Heiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Uterback T.T., Feldlyum T.V.,
RA Yang Q.Q., Haas B.J., Sun B.B., Peterson J.D., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OJ112 G08 genomic sequence."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF377947; AAM34395.2; -.

```

DR EMBL: AC135225; AAP68355.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 516 AA; 58268 MW; 0B9B6A7D0273F55 CRC64;

Query Match 1.3%; Score 8; DB 10; Length 516;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SDALDDSD 178
 |||||
 DB 307 SDALDDSD 314

RESULT 15

Q8R9U3 PRELIMINARY; PRT; 525 AA.
 AC Q8R9U3;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Predicted kinase related to dihydroxyacetone kinase.
 GN TTE1493.
 OS Thermomanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermomanaerobacteriales;
 OC Thermomanaerobacteriaceae; Thermomanaerobacter.
 OC NCBI_TaxID=119072;
 OX (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=MB4 / JCM 11007;
 RC MEDLINE=21992816; PubMed=11997336;
 RX Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome."
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE013106; AM24711.1; -
 DR GO: GO:0004371; P:glycerol kinase activity; IEA.
 DR GO: GO:0006071; P:glycerol metabolism; IEA.
 DR InterPro: IPR004007; DAK2.
 DR Pfam: PF02734; DAK2; 1.
 DR Kinase; Complete proteome.
 KW SEQUENCE 525 AA; 58638 MW; 6FB4C8215FFDEB2C CRC64;
 SQ

Query Match 1.3%; Score 8; DB 16; Length 525;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 DLKEKPE 609
 |||||
 DB 405 DLKEKPE 412

Search completed: April 6, 2004, 19:43:28
 Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:40:52 ; Search time 23 Seconds
(without alignments)
1402.880 Million cell updates/sec

Title: US-10-030-829-3
Perfect score: 625
Sequence: 1 MESSRAGPMSEKKNVQGYRP.....EFDEALQMLYKHGLNEDD 625

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.3	327	US-09-252-991A-30948	Sequence 30948, A
2	8	1.3	418	US-09-489-039A-8728	Sequence 8728, Ap
3	8	1.3	521	US-09-543-681A-5969	Sequence 5969, Ap
4	7	1.1	85	US-09-543-681A-4934	Sequence 4934, Ap
5	7	1.1	94	US-08-629-291A-31	Sequence 31, Appl
6	7	1.1	94	US-08-658-335B-31	Sequence 31, Appl
7	7	1.1	94	US-09-406-640-31	Sequence 31, Appl
8	7	1.1	102	US-09-543-681A-7823	Sequence 7823, Ap
9	7	1.1	108	US-09-252-991A-28062	Sequence 28062, A
10	7	1.1	162	US-09-252-991A-22093	Sequence 22093, A
11	7	1.1	175	US-09-705-621-13	Sequence 13, Appl
12	7	1.1	223	US-09-543-681A-4354	Sequence 4354, Ap
13	7	1.1	257	US-08-451-947-4	Sequence 4, Appl1
14	7	1.1	257	US-08-424-826A-4	Sequence 4, Appl1
15	7	1.1	257	US-08-910-691-7	Sequence 7, Appl1
16	7	1.1	257	US-08-928-694-4	Sequence 4, Appl1
17	7	1.1	257	US-08-450-842-4	Sequence 4, Appl1
18	7	1.1	257	US-08-451-390-4	Sequence 4, Appl1
19	7	1.1	257	PCT-US91-06950-4	Sequence 4, Appl1
20	7	1.1	283	US-09-434-774-6	Sequence 6, Appl1
21	7	1.1	288	US-09-813-918-3	Sequence 3, Appl1
22	7	1.1	295	US-09-134-000C-5063	Sequence 5063, Ap
23	7	1.1	300	US-09-489-039A-12934	Sequence 12934, A
24	7	1.1	312	US-09-252-991A-31204	Sequence 31204, A
25	7	1.1	351	US-08-688-988-28	Sequence 28, Appl1
26	7	1.1	361	US-09-489-039A-12349	Sequence 12349, A
27	7	1.1	375	US-09-107-532A-4312	Sequence 4312, Ap

28	7	1.1	389	US-09-489-039A-10992	Sequence 10992, A
29	7	1.1	395	US-09-673-395A-612	Sequence 612, Appl
30	7	1.1	445	US-09-252-991A-28986	Sequence 28986, A
31	7	1.1	453	US-09-711-164-349	Sequence 349, Appl
32	7	1.1	454	US-09-813-918-2	Sequence 2, Appl1
33	7	1.1	456	US-09-543-681A-5516	Sequence 5516, Ap
34	7	1.1	511	US-09-252-991A-28608	Sequence 28608, A
35	7	1.1	519	US-09-489-039A-11591	Sequence 11591, A
36	7	1.1	528	US-09-356-806-8	Sequence 8, Appl1
37	7	1.1	530	US-09-180-852-2	Sequence 2, Appl1
38	7	1.1	530	US-09-356-806-113	Sequence 113, Appl
39	7	1.1	552	US-09-252-991A-29652	Sequence 29652, A
40	7	1.1	577	US-08-756-317-13	Sequence 13, Appl1
41	7	1.1	578	US-08-981-215-1	Sequence 1, Appl1
42	7	1.1	589	US-09-543-681A-6155	Sequence 6155, Ap
43	7	1.1	596	US-09-252-991A-26463	Sequence 26463, A
44	7	1.1	630	US-09-252-991A-19822	Sequence 19822, A
45	7	1.1	633	US-09-252-991A-26576	Sequence 26576, A

ALIGNMENTS

```

RESULT 1
US-09-252-991A-30948
; Sequence 30948, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30948
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30948

Query Match      1.3%; Score 8; DB 4; Length 327;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      109 GGRGRLSR 116
Db      113 GGRGRLSR 120

RESULT 2
US-09-489-039A-8728
; Sequence 8728, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bireton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709 2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8728
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8728

```

Query Match 1.3%; Score 8; DB 4; Length 418;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 GARRVKLH 261
 DB 373 GARRVKLH 380

RESULT 3
 US-09-543-681A-5969
 ; Sequence 5969, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:

APPLICANT: GARY BRETON
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 FILE REFERENCE: 2709, 1002-001
 CURRENT APPLICATION NUMBER: US/09/543, 681A
 CURRENT FILING DATE: 2000-04-05
 PRIOR APPLICATION NUMBER: US 60/128, 706
 PRIOR FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 8344
 SEQ ID NO 5969
 LENGTH: 521
 TYPE: PRT
 ORGANISM: Proteus mirabilis

US-09-543-681A-5969

Query Match 1.3%; Score 8; DB 4; Length 521;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GGWEVTS 48
 DB 494 GGWEVTS 501

RESULT 4
 US-09-543-681A-4934
 ; Sequence 4934, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:

APPLICANT: GARY BRETON
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 FILE REFERENCE: 2709, 1002-001
 CURRENT APPLICATION NUMBER: US/09/543, 681A
 CURRENT FILING DATE: 2000-04-05
 PRIOR APPLICATION NUMBER: US 60/128, 706
 PRIOR FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 8344
 SEQ ID NO 4934
 LENGTH: 85
 TYPE: PRT
 ORGANISM: Proteus mirabilis

US-09-543-681A-4934

Query Match 1.1%; Score 7; DB 4; Length 85;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 EINSEKL 476
 DB 13 EINSEKL 19

RESULT 5
 US-08-629-291A-31
 ; Sequence 31, Application US/08629291A
 ; Patent No. 5959174
 ; GENERAL INFORMATION:

APPLICANT: Coruzzi, Gloria
 APPLICANT: Oliveira, Igor
 APPLICANT: Lam, Hon-Ming
 APPLICANT: Hsieh, Ming-Hsiun
 TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/629, 291A
 FILING DATE: 08-APR-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 5914-050
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 869-9741/8864
 TELEFAX: (212) 869-9741/8864
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 94 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-629-291A-31

Query Match 1.1%; Score 7; DB 2; Length 94;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 ONINPSS 549
 DB 64 ONINPSS 70

RESULT 6
 US-08-658-335B-31
 ; Sequence 31, Application US/08658335B
 ; Patent No. 5981703
 ; GENERAL INFORMATION:

APPLICANT: Coruzzi, Gloria
 APPLICANT: Oliveira, Igor
 APPLICANT: Lam, Hon-Ming
 APPLICANT: Hsieh, Ming-Hsiun
 TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/658, 335B

FILING DATE: 05-JUN-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 5914-052
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 94 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-658-335B-31

Query Match 1.1%; Score 7; DB 2; Length 94;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 QINPSS 549
 DB 64 QINPSS 70

RESULT 7
 US-09-406-640-31
 Sequence 31, Application US/09406640
 Patent No. 6451546

GENERAL INFORMATION:
 APPLICANT: Coruzzi, Gloria
 Oliveira, Igor
 Lam, Hon-Ming
 Hsieh, Ming-Hsiun

TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/406,640
 FILING DATE: 27-Sep-1999
 CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 5914-082
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 94 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-09-406-640-31

Query Match 1.1%; Score 7; DB 4; Length 94;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 QINPSS 549
 DB 64 QINPSS 70

RESULT 8
 US-09-543-681A-7823
 Sequence 7823, Application US/09543681A
 Patent No. 6605709

GENERAL INFORMATION:
 APPLICANT: GARY BRETON
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 FILE REFERENCE: 2709,1002-001
 CURRENT APPLICATION NUMBER: US/09/543,681A
 CURRENT FILING DATE: 2000-04-05
 PRIOR APPLICATION NUMBER: US 60/128,706
 PRIOR FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 8344
 SEQ ID NO 7823
 LENGTH: 102
 TYPE: PRT
 ORGANISM: Proteus mirabilis

US-09-543-681A-7823

Query Match 1.1%; Score 7; DB 4; Length 102;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 KIHRELA 265
 DB 59 KIHRELA 65

RESULT 9
 US-09-252-991A-28062
 Sequence 28062, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196,136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 28062
 LENGTH: 108
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-28062

Query Match 1.1%; Score 7; DB 4; Length 108;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ARGSAQ 147
 DB 42 ARGSAQ 48

RESULT 10
 US-09-252-991A-22093
 Sequence 22093, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22093
LENGTH: 162
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22093

Query Match 1.1%; Score 7; DB 4; Length 162;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 VVGQOQO 543
DB 141 VVGQOQO 147

RESULT 11
US-09-705-621-13
Sequence 13, Application US/09705621
Patent No. 667335
GENERAL INFORMATION:
APPLICANT: ESTES, MARY
TITLE OF INVENTION: ROTAVIRUS ENTEROTOXIN NSP4 AND METHODS OF USING SAME
FILE REFERENCE: P01932US3
CURRENT APPLICATION NUMBER: US/09/705,621
CURRENT FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent version 3.0
SEQ ID NO 13
LENGTH: 175
TYPE: PRT
ORGANISM: Rotavirus
US-09-705-621-13

Query Match 1.1%; Score 7; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 VVKELRQ 440
DB 102 VVKELRQ 108

RESULT 12
US-09-543-681A-4354
Sequence 4354, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4354
LENGTH: 233
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-4354

Query Match 1.1%; Score 7; DB 4; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 KLSKQNK 460
DB 162 KLSKQNK 168

RESULT 13
US-08-451-947-4
Sequence 4, Application US/08451947
Patent No. 5702906
GENERAL INFORMATION:
APPLICANT: GENENTECH, INC.
TITLE OF INVENTION: NOVEL NEUROTROPHIC FACTOR
NUMBER OF SEQUENCES: 100
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,947
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/426419
FILING DATE: 19-APR-1995
PRIOR APPLICATION DATA: 08/030013
APPLICATION NUMBER: 22-MAR-1993
PRIOR APPLICATION DATA: 07/648482
FILING DATE: 31-JAN
APPLICATION NUMBER: 07/587707
PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 666P2C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/952-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-451-947-4

Query Match 1.1%; Score 7; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KXKLSQ 458
DB 45 KXKLSQ 51

RESULT 14
US-08-424-826A-4

; Sequence 4, Application US/08424826A
; Patent No. 5830858
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Arnon
; TITLE OF INVENTION: NOVEL NEUTROPHIC FACTOR
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,826A
; FILING DATE: 19-Apr-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/240387
; FILING DATE: 10-May-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/648482
; FILING DATE: 31-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/587707
; FILING DATE: 25-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, PhD., Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: P0666P1C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 910/371-7168
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-424-826A-4

Query Match 1.1%; Score 7; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
DB 45 KNTLSKQ 51

RESULT 15
US-08-910-691-7
; Sequence 7, Application US/08910691
; Patent No. 6015552
; GENERAL INFORMATION:
; APPLICANT: WATANABE, Tatsuya
; APPLICANT: YOSHITOMI, Sumie
; APPLICANT: SASADA, Reiko
; TITLE OF INVENTION: THERAPEUTIC AGENT FOR NEUTROPENIA
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,691
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,969
; FILING DATE: 19930604
; ATTORNEY/AGENT INFORMATION:
; NAME: NEUNER, George W
; REGISTRATION NUMBER: 26964
; REFERENCE/DOCKET NUMBER: 12345
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELETYPE: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-910-691-7

Query Match 1.1%; Score 7; DB 3; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
DB 45 KNTLSKQ 51

Search completed: April 6, 2004, 19:44:36
Job time : 24 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:43:32 ; Search time 46 Seconds
(without alignments)
3568.388 Million cell updates/sec

Title: US-10-030-829-3
Perfect score: 625
Sequence: 1 MSRRAGPMSEKKNVQGGYRP.....EPFDEALQMLYKHLNEDD 625

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 1071772 seqs, 26263353 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/PC7US_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.4	223	US-10-424-599-234019	Sequence 234019,
2	9	1.4	227	US-09-738-626-5288	Sequence 5288, Ap
3	9	1.4	324	US-10-424-599-234018	Sequence 234018,
4	9	1.4	382	US-10-369-493-12141	Sequence 12141, A
5	9	1.4	611	US-10-425-114-58390	Sequence 58390, A
6	8	1.3	916	US-10-282-122A-51413	Sequence 51413, A
7	7	1.1	18	US-10-225-567A-1514	Sequence 1514, Ap
8	7	1.1	36	US-09-205-658-225	Sequence 225, App
9	7	1.1	36	US-09-963-693-225	Sequence 250491,
10	7	1.1	56	US-10-424-599-250491	Sequence 31679, A
11	7	1.1	56	US-10-029-386-31679	Sequence 38645, Ap
12	7	1.1	57	US-10-264-049-3845	Sequence 154923,
13	7	1.1	60	US-10-424-599-154923	Sequence 228124,
14	7	1.1	62	US-10-424-599-228124	Sequence 169910,
15	7	1.1	68	US-10-424-599-169910	

16	7	1.1	69	US-10-106-698-5320	Sequence 5320, Ap
17	7	1.1	78	US-10-424-599-280255	Sequence 280255,
18	7	1.1	94	US-10-223-047-31	Sequence 31, Appl
19	7	1.1	104	US-09-764-869-650	Sequence 650, App
20	7	1.1	104	US-10-091-504-650	Sequence 650, App
21	7	1.1	104	US-10-227-577-650	Sequence 650, App
22	7	1.1	121	US-10-108-260A-4862	Sequence 4862, Ap
23	7	1.1	136	US-10-425-114-56375	Sequence 56375, A
24	7	1.1	136	US-10-425-114-69428	Sequence 69428, A
25	7	1.1	138	US-10-424-599-177015	Sequence 177015,
26	7	1.1	138	US-10-155-886-53	Sequence 53, Appl
27	7	1.1	138	US-10-155-886-59	Sequence 59, Appl
28	7	1.1	139	US-10-155-886-55	Sequence 57, Appl
29	7	1.1	139	US-10-155-886-57	Sequence 51842, A
30	7	1.1	141	US-10-282-122A-51842	Sequence 22038, A
31	7	1.1	148	US-10-369-493-22038	Sequence 72748, A
32	7	1.1	166	US-10-425-114-72748	Sequence 277920,
33	7	1.1	173	US-10-424-599-277920	Sequence 54097, A
34	7	1.1	180	US-10-425-114-54097	Sequence 283874,
35	7	1.1	189	US-10-424-599-283874	Sequence 64832, A
36	7	1.1	196	US-10-425-114-64832	Sequence 128, App
37	7	1.1	211	US-10-627-476-128	Sequence 26, Appl
38	7	1.1	216	US-09-861-451A-26	Sequence 69066, A
39	7	1.1	219	US-10-282-122A-69066	Sequence 5665, Ap
40	7	1.1	226	US-10-335-977-5665	Sequence 32003, A
41	7	1.1	227	US-10-335-977-5665	Sequence 4, Appl
42	7	1.1	253	US-10-029-386-32003	Sequence 5, Appl
43	7	1.1	257	US-08-450-842-4	
44	7	1.1	257	US-09-788-186-5	
45	7	1.1	257	US-09-788-186-6	

ALIGNMENTS

RESULT 1
US-10-424-599-234019
; Sequence 234019, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424, 599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 285684
; SEQ. ID NO 234019
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(223)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1.pep
; US-10-424-599-234019

Query Match 1.4%; Score 9; DB 12; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 RQWHPACQ 232
Db 119 RQWHPACQ 127

RESULT 2
US-09-738-626-5288

```

; Sequence 5288, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIYOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5288
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5288

```

```

Query Match      1.4%; Score 9; DB 9; Length 227;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      170 DSDALDSD 178
        |||||
DB      179 DSDALDSD 187

```

```

RESULT 3
US-10-424-599-234018
; Sequence 234018, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234018
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1.pep
US-10-424-599-234018

```

```

Query Match      1.4%; Score 9; DB 12; Length 324;
Best Local Similarity 100.0%; Pred. No. 5;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      224 RQMHCPACQ 232
        |||||
DB      119 RQMHCPACQ 127

```

RESULT 4

```

US-10-369-493-12141
; Sequence 12141, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12141
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12141

```

```

Query Match      1.4%; Score 9; DB 15; Length 382;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      261 HRELAETVE 269
        |||||
DB      16 HRELAETVE 24

```

```

RESULT 5
US-10-425-114-58390
; Sequence 58390, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58390
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3079-030-A8_F11.pep
US-10-425-114-58390

```

```

Query Match      1.4%; Score 9; DB 12; Length 611;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      233 NGPGADMY 241
        |||||
DB      218 NGPGADMY 226

```

```

RESULT 6
US-10-282-122A-51413
; Sequence 51413, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos

```



```
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Cair, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 51413
LENGTH: 916
TYPE: PRT
ORGANISM: Bordetella pertussis
US-10-282-122A-51413

Query Match
Best Local Similarity 1.3%; Score 8; DB 12; Length 916;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 FDEALEQL 614
Db 654 FDEALEQL 661

RESULT 7
US-10-225-567A-1514
; Sequence 1514, Application US/10225567A
; Publication No. US20030113798A1
GENERAL INFORMATION:
APPLICANT: Lifespan Biosciences
APPLICANT: Brown, Joseph P.
APPLICANT: Burmer, Glenna C.
APPLICANT: Roush, Christine L.
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 1920-4-4
CURRENT APPLICATION NUMBER: US/10/225,567A
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/257,144
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 2292
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1514
LENGTH: 18
TYPE: PRT
ORGANISM: Homo sapiens
```

```
US-10-225-567A-1514

Query Match
Best Local Similarity 1.1%; Score 7; DB 14; Length 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 DSDDDLA 182
Db 3 DSDDDLA 9

RESULT 8
US-09-205-658-225
; Sequence 225, Application US/09205658
; Patent No. US20010029617A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Oeg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/205,658
CURRENT FILING DATE: 1998-12-03
EARLIER APPLICATION NUMBER: 08/857,076
EARLIER FILING DATE: 1997-05-15
EARLIER APPLICATION NUMBER: 08/888,534
EARLIER FILING DATE: 1997-07-07
EARLIER APPLICATION NUMBER: US98/10080
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 225
LENGTH: 36
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-205-658-225

Query Match
Best Local Similarity 1.1%; Score 7; DB 9; Length 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 LRQISED 444
Db 2 LRQISED 8

RESULT 9
US-09-963-693-225
; Sequence 225, Application US/09963693
; Publication No. US20030181364A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Oeg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/963,693
CURRENT FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US/09/205,658
PRIOR FILING DATE: 1998-12-03
PRIOR APPLICATION NUMBER: 08/857,076
PRIOR FILING DATE: 1997-05-15
PRIOR APPLICATION NUMBER: 08/888,534
PRIOR FILING DATE: 1997-07-07
PRIOR APPLICATION NUMBER: US98/10080
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 225
LENGTH: 36
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-963-693-225
```

Query Match 1.1%; Score 7; DB 10; Length 36;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 LRQISED 444
 DB 2 LRQISED 8

RESULT 10

US-10-424-599-250491
 ; Sequence 250491, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 250491
 ; LENGTH: 56
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_68221C.1.pcp
 ; US-10-424-599-250491

Query Match 1.1%; Score 7; DB 12; Length 56;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 ASYIPCG 283
 DB 31 ASYIPCG 37

RESULT 11

US-10-029-386-31679
 ; Sequence 31679, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME ANALYSIS TWO
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 31679
 ; LENGTH: 56
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC008974.6
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.87
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.55
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
 ; OTHER INFORMATION: SWISSPROT HIT: Q00535, EVALU8 2.80e+00
 ; US-10-029-386-31679

Query Match 1.1%; Score 7; DB 14; Length 56;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHREL 376
 DB 47 ERLHREL 53

RESULT 12

US-10-264-049-3845
 ; Sequence 3845, Application US/10264049
 ; Publication No. US20040005579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA133PI
 ; CURRENT APPLICATION NUMBER: US/10/264,049
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: US 60/209,467
 ; PRIOR FILING DATE: 2000-06-07
 ; NUMBER OF SEQ ID NOS: 4360
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 3845
 ; LENGTH: 57
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (1)
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
 ; US-10-264-049-3845

Query Match 1.1%; Score 7; DB 15; Length 57;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 QNINPSS 549
 DB 27 QNINPSS 33

RESULT 13

US-10-424-599-154923
 ; Sequence 154923, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 154923
 ; LENGTH: 60
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_110917C.1.pcp
 ; US-10-424-599-154923

Query Match 1.1%; Score 7; DB 12; Length 60;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 EYEOLVQ 27
 DB 47 EYEOLVQ 53

RESULT 14

```
US-10-424-599-228124
; Sequence 228124, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 228124
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(62)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48023C.1.pep
; US-10-424-599-228124

Query Match          1.1%; Score 7; DB 12; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      350 GHRGMSV 356
Db       30 GHRGMSV 36

RESULT 15
US-10-424-599-169910
; Sequence 169910, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169910
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124443C.1.pep
; US-10-424-599-169910

Query Match          1.1%; Score 7; DB 12; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      561 SSFIEFQ 567
Db       22 SSFIEFQ 28
```

Search completed: April 6, 2004, 19:49:22
Job time : 47 secs

This Page Blank (uspio)